Stochastic Processes and Finance
(part one)

Dr Nic Freeman

May 8, 2017

Contents

0 Introduction **3**

0.1 Syllabus . . . 4
0.2 Problem sheets . . . 4
0.3 Examination . . . 4
0.4 Website . . . 4

1 Expectation and Arbitrage **5**

1.1 Betting on coin tosses . . . 5
1.2 The one-period market . . . 6
1.3 Arbitrage . . . 8
1.4 Modelling discussion . . . 11
1.5 Exercises . . . 12

2 Probability spaces and random variables **13**

2.1 Probability measures and σ-fields . . . 13
2.2 Random variables . . . 16
2.3 Two kinds of examples . . . 20
2.4 Expectation . . . 23
2.5 Exercises . . . 26

3 Conditional expectation and martingales **27**

3.1 Conditional expectation . . . 27
3.2 Properties of conditional expectation . . . 31
3.3 Martingales . . . 33
3.4 Exercises . . . 36

4 Stochastic processes **37**

4.1 Random walks . . . 37
4.2 Urn processes . . . 40
4.3 A branching process . . . 42
4.4 Other stochastic processes . . . 44
4.5 Exercises . . . 45

5 The binomial model **46**

5.1 Arbitrage in the one-period model . . . 46
5.2 Hedging in the one-period model . . . 50

5.3 Types of financial derivative . . . 54
5.4 The binomial model . . . 55
5.5 Portfolios, arbitrage and martingales . . . 56
5.6 Hedging . . . 59
5.7 Exercises . . . 63

6 Convergence of random variables 64
6.1 Modes of convergence . . . 64
6.2 The dominated convergence theorem . . . 66
6.3 Exercises . . . 67

7 Stochastic processes and martingale theory 68
7.1 The martingale transform . . . 68
7.2 Roulette . . . 70
7.3 The optional stopping theorem . . . 71
7.4 Hitting times of random walks . . . 73
7.5 Exercises . . . 78

8 Further theory of stochastic processes (Δ) 79
8.1 The Borel-Cantelli lemmas (Δ) . . . 79
8.2 The martingale convergence theorem (Δ) . . . 81
8.3 Long term behaviour of Galton-Watson processes (Δ) . . . 83
8.4 Exercises (Δ) . . . 84

A Solutions to exercises 85

Chapter 0

Introduction

We live in a random world; we cannot be certain of tomorrows weather or what the price of petrol will be next year. But randomness is never 'completely' random. Often we know, or rather, believe that some events are likely and others are unlikely. We might think that two events are both possible, but are unlikely to occur together, and so on.

How should we handle this situation? Naturally, we would like to understand the world around us and, when possible, to anticipate what might happen in the future. This necessitates that we study the variety of random processes that we find around us.

We will see many and varied examples of random processes throughout this course, although we will tend to call them *stochastic processes* (with the same meaning). They reflect the wide variety of unpredictable ways in which reality behaves. We will also introduce a key idea used in the study of stochastic processes, known as a martingale.

It has become common, in both science and industry, to use highly complex models of the world around us. Such models cannot be magicked out of thin air. In fact, in much the same way as we might build a miniature space station out of individual pieces of Lego, what is required is a set of useful pieces that can be fitted together into realistic models. The theory of stochastic processes provides some of the most useful pieces, and the models built from them are generally called stochastic models.

One industry that makes extensive use of stochastic modelling is *finance*. In this course, we will often use financial models to motivate our discussion of stochastic processes.

The central question in a financial model is usually how much a particular object is worth. For example, we might ask how much we need to pay today, to have a barrel of oil delivered in six months time. We might ask for something more complicated: how much would it cost to have the opportunity, in six months time, to buy a barrel of oil, for a price that is agreed on now? We will study the Black-Scholes model and the concept of 'arbitrage free pricing', which provide somewhat surprising answers to this type of question.

0.1 Syllabus

These notes are for three courses: MAS352, MAS452 and MAS6052.

Some sections of the course are included in MAS452/6052 but not in MAS352. These sections are marked with a (Δ) symbol. We will not cover these sections in lectures. Students taking MAS452/6052 should study these sections independently.

A small number of remarks, proofs etc are marked with a ($\star$) symbol. These are often cases where connections are made to and from other areas of mathematics; they are off-syllabus but still important to understand. We will cover these parts in lectures.

0.2 Problem sheets

The exercises are divided up according to the chapters of the course. Exercises marked with a (Δ) are intended only for MAS452/6052 students.

It is expected that students will attempt all exercises (for the version of the course they are taking) and review their own solutions using the typed solutions provided at the end of these notes. Some exercises are marked as 'challenge questions' – these are typically difficult questions designed to test ingenuity.

At three points during each semester, a selection of exercises will set for handing in. These will be marked and returned in lectures.

0.3 Examination

The whole course will be examined in the summer sitting. Parts of the course marked with a (Δ) are examinable for MAS462/6052 but not for MAS352.

Parts of the course marked with a ($\star$) will not be examined (for everyone).

0.4 Website

Further information, including the timetable, can be found on `http://nicfreeman.staff.shef.ac.uk/MASx52/`.

Chapter 1

Expectation and Arbitrage

In this chapter we look at our first example of a financial market. We introduce the idea of arbitrage free pricing, and discuss what tools we would need to build better models.

1.1 Betting on coin tosses

We begin by looking at a simple betting game. Someone tosses a fair coin. They offer to pay you \$1 if the coin comes up heads and nothing if the coin comes up tails. How much are you prepared to pay to play the game?

One way that you might answer this question is to look at the expected return of playing the game. If the (random) amount of money that you win is X, then you'd expect to make

$$\mathbb{E}[X] = \tfrac{1}{2}\$1 + \tfrac{1}{2}\$0 = \$0.50.$$

So you might offer to pay \$0.50 to play the game.

We can think of a single play as us paying some amount to buy *a random quantity*. That is, we pay \$0.50 to buy the random quantity X, then later on we discover if X is \$1 or \$0.

We can link this 'pricing by expectation' to the long term average of our winnings, if we played the game multiple times. Formally this uses the strong law of large numbers:

Theorem 1.1.1 *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of random variables that are independent and identically distributed. Suppose that $\mathbb{E}[X_1]=\mu$ and $\mathrm{var}(X_1)<\infty$, and set*

$$S_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Then, with probability one, $S_n \to \mu$ as $n \to \infty$.

In our case, if we played the game a large number n of times, and on play i our winnings were X_i, then our average winnings would be $S_n \approx \mathbb{E}[X_1] = \frac{1}{2}$. So we might regard \$0.50 as a 'fair' price to pay for a single play. If we paid less, in the long run we'd make money, and if we paid more, in the long run we'd lose money.

Often, though, you might not be willing to pay this price. Suppose your life savings were \$20,000. You probably (hopefully) wouldn't gamble it on the toss of a single coin, where you would get \$40,000 on heads and \$0 on tails; it's too risky.

It is tempting to hope that the fairest way to price anything is to calculate its expected value, and then charge that much. As we will explain in the rest of Chapter 1, this tempting idea turns out to be *completely* wrong.

1.2 The one-period market

Let's replace our betting game by a more realistic situation. This will require us to define some terminology. Our convention, for the whole of the course, is that when we introduce a new piece of financial terminology we'll write it in **bold**.

A **market** is any setting where it is possible to buy and sell one or more quantities. An object that can be bought and sold is called a **commodity**. For our purposes, we will always define exactly what can be bought and sold, and how the value of each commodity changes with time. We use the variable t for time.

It is important to realize that money itself is a commodity. It can be bought and sold, in the sense that it is common to exchange money for some other commodity. For example, we might exchange some money for a new pair of shoes; at that same instant someone else is exchanging a pair of shoes for money. When we think of money as a commodity we will usually refer to it as **cash** or as a cash bond.

In this section we define a market, which we'll then study for the rest of Chapter 1. It will be a simple market, with only two commodities. Naturally, we have plans to study more sophisticated examples, but we should start small!

Unlike our coin toss, in our market we will have *time*. As time passes money earns interest, or if it is money that we owe we will be required to pay interest. We'll have just one step of time in our simple market. That is, we'll have time $t = 0$ and time $t = 1$. For this reason, we will call our market the **one-period market**.

Let $r > 0$ be a deterministic constant, known as the **interest rate**. If we put an amount x of cash into the bank at time $t = 0$ and leave it there until time $t = 1$, the bank will then contain

$$x(1 + r)$$

in cash. The same formula applies if x is negative. This corresponds to borrowing C_0 from the bank (i.e. taking out a loan) and the bank then requires us to pay interest on the loan.

Our market contains cash, as its first commodity. As its second, we will have a **stock**. Let us take a brief detour and explain what is meant by stock.

Firstly, we should realize that companies can be (and frequently are) owned by more than one person at any given time. Secondly, the 'right of ownership of a company' can be broken down into several different rights, such as:

- The right to a share of the profits.
- The right to vote on decisions concerning the companies future – for example, on a possible merge with another company.

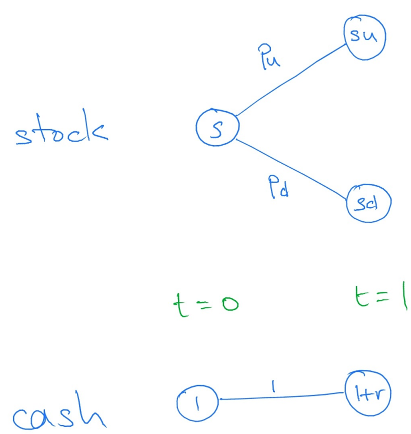
A **share** is a proportion of the rights of ownership of a company; for example a company might split its rights of ownership into 100 equal shares, which can then be bought and sold individually. The value of a share will vary over time, often according to how the successful the company is. A collection of shares in a company is known as stock.

We allow the amount of stock that we own to be any real number. This means we can own a fractional amount of stock, or even a negative amount of stock. This is realistic: in the same way as we could borrow can from a bank, we can borrow stock from a stockbroker! We don't pay any interest on borrowed stock, we just have to eventually return it. (In reality the stockbroker would charge us a fee but we'll pretend they don't, for simplicity.)

The **value** or **worth** of a stock (or, indeed any commodity) is the amount of cash required to buy a single unit of stock. This changes, randomly:

> Let $u > d > 0$ and $s > 0$ be deterministic constants. At time $t = 0$, it costs $S_0 = s$ cash to buy one unit of stock. At time $t = 1$, one unit of stock becomes worth
>
> $$S_1 = \begin{cases} su & \text{with probability } p_u, \\ sd & \text{with probability } p_d. \end{cases}$$
>
> of cash. Here, $p_u, p_d > 0$ and $p_u + p_d = 1$.

We can represent the changes in value of cash and stocks as a tree, where each edge is labelled by the probability of occurrence.



To sum up, in the one-period market it is possible to trade stocks and cash. There are two points in time, $t = 0$ and $t = 1$.

- It we have x units of cash at time $t = 0$, they will become worth $x(1 + r)$ at time $t = 1$.
- If we have y units of stock, that are worth $yS_0 = sy$ at time $t = 0$, they will become worth

  $$yS_1 = \begin{cases} ysu & \text{with probability } p_u, \\ ysd & \text{with probability } p_d. \end{cases}$$

  at time $t = 1$.

We place no limit on how much, or how little, of each can be traded. That is, we assume the bank will loan/save as much cash as we want, and that we are able to buy/sell unlimited amounts of stock at the current market price. A market that satisfies this assumption is known as **liquid** market.

For example, suppose that $r > 0$ and $s > 0$ are given, and that $u = \frac{3}{2}$, $d = \frac{1}{3}$ and $p_u = p_d = \frac{1}{2}$. At time $t = 0$ we hold a portfolio of 5 cash and 8 stock. What is the expected value of this portfolio at time $t = 1$?

Our 5 units of cash become worth $5(1+r)$ at time 1. Our 8 units of stock, which are worth $8S_0$ at time 0, become worth $8S_1$ at time 1. So, at time $t=1$ our portfolio is worth $V_1 = 5(1+r) + 8S_1$ and the expected value of our portfolio and time t is

$$\begin{aligned}\mathbb{E}[V_1] &= 5(1+r) + 8sup_u + 8sdp_d \\ &= 5(1+r) + 6s + \frac{4}{3}s \\ &= 5 + 5r + \frac{22}{3}s.\end{aligned}$$

1.3 Arbitrage

We now introduce a key concept in mathematical finance, known as **arbitrage**. We say that arbitrage occurs in a market if it possible to make money, for free, without risk of losing money.

There is a subtle distinction to be made here. We might sometimes *expect* to make money, on average. But an arbitrage possibility only occurs when it is possible to make money without any chance of losing it.

Example 1.3.1 *Suppose that, in the one-period market, someone offered to sell us a single unit of stock for a special price* $\frac{s}{2}$ *at time* 0. *We could then:*

1. *Take out a loan of* $\frac{s}{2}$ *from the bank.*
2. *Buy the stock, at the special price, for* $\frac{s}{2}$ *cash.*
3. *Sell the stock, at the market rate, for* s *cash.*
4. *Repay our loan of* $\frac{s}{2}$ *to the bank (we still are at* $t=0$*, so no interest is due).*
5. *Profit!*

We now have no debts and $\frac{s}{2}$ *cash, with certainty. This is an example of arbitrage.*

Example 1.3.1 is obviously artificial. It does illustrates an important point: no one should sell anything at a price that makes an arbitrage possible. However, if nothing is sold at a price that would permit arbitrage then, equally, nothing can be bought for a price that would permit arbitrage. With this in mind:

> We assume that no arbitrage can occur in our market.

Let us step back and ask a natural question, about our market. Suppose we wish to have a single unit of stock delivered to us at time T, but we want to agree in advance, at time 0, what price K we will pay for it. To do so, we would enter into a **contract**. A contract is an agreement between two (or more) parties (i.e. people, companies, institutions, etc) that they will do something together.

Consider a contract that refers to one party as the buyer and another party as the seller. The contract specifies that:

> At time 1, the seller will be paid K cash and will deliver one unit of stock to the buyer.

A contract of this form is known as a **forward** contract. Note that no money changes hands at time 0. The price K that is paid at time 1 is known as the **strike price**. The question is: what should be the value of K?

In fact, there is *only one* possible value for K. This value is

$$K = s(1 + r). \tag{1.1}$$

Let us now explain why. We argue by contradiction.

- Suppose that a price $K > s(1 + r)$ was agreed. Then we could do the following:

  1. At time 0, enter into a forward contract as the seller.
  2. Borrow s from the bank, and use it buy a single unit of stock.
  3. Wait until time 1.
  4. Sell the stock (as agreed in our contract) in return for K cash.
  5. We owe the bank $s(1 + r)$ to pay back our loan, so we pay this amount to the bank.
  6. We are left with $K - s(1 + r) > 0$ profit, in cash.

  With this strategy we are *certain* to make a profit. This is arbitrage!

- Suppose, instead, that a price $K < s(1 + r)$ was agreed. Then we could:

  1. At time 0, enter into a forward contract as the buyer.
  2. Borrow a single unit of stock from the stockbroker.
  3. Sell this stock, in return for s cash.
  4. Wait until time 1.
  5. We now have $s(1 + r)$ in cash. Since $K < s(1 + r)$ we can use K of this cash to buy a single unit of stock (as agreed in our contract).
  6. Use the stock we just bought to pay back the stockbroker.
  7. We are left with $s(1 + r) - K > 0$ profit, in cash.

  Once again, with this strategy we are *certain* to make a profit. Arbitrage!

Therefore, we reach the surprising conclusion that the only possible choice is $K = s(1 + r)$. We refer to $s(1 + r)$ as the arbitrage free value for K. This is our first example of an important principle:

> The absence of arbitrage can force prices to take particular values. This is known as **arbitrage free pricing**.

1.3.1 Expectation versus arbitrage

What of pricing by expectation? Let us, temporarily, forget about arbitrage and try to use pricing by expectation to find K.

The value of our forward contract at time 1, from the point of view of the buyer, is $S_1 - K$. It costs nothing to enter into the forward contract, so if we believed that we should price

Figure 1.1: The stock price in GBP of Lloyds Banking Group, from September 2011 to September 2016.

the contract by its expectation, we would want it to cost nothing! This would mean that $\mathbb{E}[S_1 - K] = 0$, which means we'd choose

$$K = \mathbb{E}[S_1] = sup_u + sdp_d. \tag{1.2}$$

This is *not* the same as the formula $K = s(1+r)$, which we deduced in the previous section.

It is possible that our two candidate formulas for K 'accidentally' agree, that is $up_u + dp_d = s(1+r)$, but they only agree for very specific values of u, p_u, d, p_d, s and r. Observations of real markets show that this doesn't happen.

It may feel very surprising that (1.2) is different to (1.1). The reality is, that financial markets are arbitrage free, and the *correct* strike price for our forward contract is $K = s(1+r)$. However intuitively appealing it might seem to price by expected value, it is not what happens in reality.

Does this mean that, with the correct strike price $K = s(1+r)$, on average we either make or lose money by entering into a forward contract? Yes, it does. But investors are often not concerned with average payoffs – the world changes too quickly to make use of them. Investors are concerned with what happens to them *personally*. Having realized this, we can give a short explanation, in economic terms, of why markets are arbitrage free.

If it is possible to carry out arbitrage within a market, traders will[1] discover how and immediately do so. This creates high demand to buy undervalued commodities. It also creates high demand to borrow overvalued commodities. In turn, this demand causes the price of commodities to adjust, until it is no longer possible to carry out arbitrage. The result is that the market constantly adjusts, and stays in an equilibrium in which no arbitrage is possible.

Of course, in many respects our market is an imperfect model. We will discuss its shortcomings, as well as produce better models, as part of the course.

Remark 1.3.2 *We will not mention 'pricing by expectation' again in the course. In a liquid market, arbitrage free pricing is what matters.*

[1] Usually.

1.4 Modelling discussion

Our proof that the arbitrage free value for K was $s(1+r)$ is mathematically correct, but it is not ideal. We relied on discovering specific trading strategies that (eventually) resulted in arbitrage. If we tried to price a more complicated contract, we might fail to find the right trading strategies and hence fail to find the right prices. In real markets, trading complicated contracts is common.

Happily, this is precisely the type of situation where mathematics can help. What is needed is a *systematic* way of calculating arbitrage free prices, that *always* works. In order to find one, we'll need to first develop several key concepts from probability theory. More precisely:

- **We need to be able to express the idea that, as time passes, we gain information.**

  For example, in our market, at time $t = 0$ we don't know how the stock price will change. But at time $t = 1$, it has already changed and we do know. Of course, real markets have more than one time step, and we only gain information gradually.

- **We need stochastic processes.**

  Our stock price process $S_0 \mapsto S_1$, with its two branches, is too simplistic. Real stock prices have a 'jagged' appearance (see Figure 1.1). What we need is a library of useful stochastic processes, to build models out of.

In fact, these two requirements are common to almost all stochastic modelling. For this reason, in the next two chapters (and several later chapters, too) we'll develop our probabilistic tools based on a wide range of examples. We'll return to financial markets in Chapter 5.

1.5 Exercises

On the one-period market

All these questions refer to the market defined in Section 1.2 and use notation u, d, p_u, p_d, r, s from that section.

1.1 Suppose that our portfolio at time 0 has 10 units of cash and 5 units of stock. What is the value of this portfolio at time 1?

1.2 Suppose that $0 < d < 1 + r < u$. Our portfolio at time 0 has $x \geq 0$ units of cash and $y \geq 0$ units of stock, but we will have a debt to pay at time 1 of $K > 0$ units of cash.

(a) Assuming that we don't buy or sell anything at time 0, under what conditions on x, y, K can we be certain of paying off our debt?

(b) Suppose that do allow ourselves to trade cash and stocks at time 0. What strategy gives us the best chance of being able to pay off our debt?

1.3 (a) Suppose that $0 < 1 + r < d < u$. Find a trading strategy that results in an arbitrage.

(b) Suppose instead that $0 < d < u < 1 + r$. Find a trading strategy that results in an arbitrage.

Revision of probability and analysis

1.4 Let Y be an exponential random variable with parameter $\lambda > 0$. That is, the probability density function of Y is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Calculate $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$. Hence, show that $\mathrm{var}(X) = \frac{1}{\lambda^2}$.

1.5 Let (X_n) be a sequence of independent random variables such that

$$\mathbb{P}[X_n = x] = \begin{cases} \frac{1}{n} & \text{if } x = n^2 \\ 1 - \frac{1}{n} & \text{if } x = 0. \end{cases}$$

Show that $\mathbb{P}[|X_n| > 0] \to 0$ and $\mathbb{E}[X_n] \to \infty$, as $n \to \infty$.

1.6 Let X be a normal random variable with mean μ and variance $\sigma^2 > 0$. By calculating $\mathbb{P}[Y \leq y]$ (or otherwise) show that $Y = \frac{X-\mu}{\sigma}$ is a normal random variable with mean 0 and variance 1.

1.7 For which values of $p \in (0, \infty)$ is $\int_1^\infty x^{-p}\, dx$ finite?

1.8 Which of the following sequences converge as $n \to \infty$? What do they converge too?

$$e^{-n} \qquad \sin\left(\frac{n\pi}{2}\right) \qquad \frac{\cos(n\pi)}{n} \qquad \sum_{i=1}^{n} 2^{-i} \qquad \sum_{i=1}^{n} \frac{1}{i}.$$

Give brief reasons for your answers.

1.9 Let (x_n) be a sequence of real numbers such that $\lim_{n\to\infty} x_n = 0$. Show that (x_n) has a subsequence (x_{n_r}) such that $\sum_{r=1}^{\infty} |x_{n_r}| < \infty$.

Chapter 2

Probability spaces and random variables

In this chapter we review probability theory, and develop some key tools for use in later chapters. We begin with a special focus on σ-fields. The role of a σ-field is to provide a way of controlling which information is visible (or, currently of interest) to us. As such, σ-fields will allow us to express the idea that, as time passes, we gain information.

2.1 Probability measures and σ-fields

Let Ω be a set. In probability theory, the symbol Ω is typically (and always, in this course) used to denote the *sample space*. Intuitively, we think of ourselves as conducting some random experiment, with an unknown outcome. The set Ω contains an $\omega \in \Omega$ for every possible outcome of the experiment.

Subsets of Ω correspond to collections of possible outcomes; such a subset is referred as an event. For instance, if we roll a dice we might take $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the set $\{1, 3, 5\}$ is the event that our dice roll is an odd number.

Definition 2.1.1 *Let $\mathcal{F}$ be a set of subsets of Ω. We say $\mathcal{F}$ is a σ-field if it satisfies the following properties:*

1. *$\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.*
2. *if $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$.*
3. *if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

The role of a σ-field is to choose which subsets of outcomes we are actually interested in. The power set $\mathcal{F} = \mathcal{P}(\Omega)$ is always a σ-field, and in this case every subset of Ω is an event. But $\mathcal{P}(\Omega)$ can be very big, and if our experiment is complicated, with many or even infinitely many possible outcomes, we might want to consider a smaller choice of $\mathcal{F}$ instead.

Sometimes we will need to deal with more than one σ-field at a time. A σ-field $\mathcal{G}$ such that $\mathcal{G} \subseteq \mathcal{F}$ is known as a *sub-σ-field* of $\mathcal{F}$.

We say that a subset $A \subseteq \Omega$ is *measurable*, or that it is an *event* (or *measurable event*), if $A \in \mathcal{F}$. To make to it clear which σ-field we mean to use in this definition, we sometimes write that an event is $\mathcal{F}$-measurable.

Example 2.1.2 *Some examples of experiments and the σ-fields we might choose for them are the following:*

- *We toss a coin, which might result in heads H or tails T. We take $\Omega = \{H, T\}$ and $\mathcal{F} = \big\{\Omega, \{H\}, \{T\}, \emptyset\big\}$ to be the power set of Ω.*
- *We toss two coins, both of which might result in heads H or tails T. We take $\Omega = \{HH, TT, HT, TH\}$. However, we are only interested in the outcome that both coins are heads. We take $\mathcal{F} = \big\{\Omega, \Omega \setminus \{HH\}, \{HH\}, \emptyset\big\}$.*

There are natural ways to choose a σ-field, even if we think of Ω as just an arbitrary set. For example, $\mathcal{F} = \{\Omega, \emptyset\}$ is a σ-field. If A is a subset of Ω, then $\mathcal{F} = \{\Omega, A, \Omega \setminus A, \emptyset\}$ is a σ-field (check it!).

Given Ω and $\mathcal{F}$, the final ingredient of a probability space is a measure $\mathbb{P}$, which tells us how likely the events in $\mathcal{F}$ are to occur.

Definition 2.1.3 *A probability measure $\mathbb{P}$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying:*

1. $\mathbb{P}[\Omega] = 1$.
2. *If $A_1, A_2, \ldots \in \mathcal{F}$ are pair-wise disjoint (i.e. $A_i \cap A_j = \emptyset$ for all i, j such that $i \neq j$) then*

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

The second of these conditions if often called σ-additivity. Note that we needed Definition 2.1.1 to make sense of Definion 2.1.3, because we needed something to tell us that $\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right]$ was defined!

Definition 2.1.4 *A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ-field and $\mathbb{P}$ is a probability measure.*

For example, to model a single fair coin toss we would take $\Omega = \{H, T\}$, $\mathcal{F} = \{\Omega, \{H\}, \{T\}, \emptyset\}$ and define $\mathbb{P}[H] = \mathbb{P}[T] = \frac{1}{2}$.

We commented above that often we want to choose $\mathcal{F}$ to be smaller than $\mathcal{P}(\Omega)$, but we have not yet shown how to choose a suitably small $\mathcal{F}$. Fortunately, there is a general way of doing so, for which we need the following technical lemma.

Lemma 2.1.5 *Let I be any set and for each $i \in I$ let $\mathcal{F}_i$ be a σ-field. Then*

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i \tag{2.1}$$

is a σ-field

PROOF: We check the three conditions of Definition 2.1.1 for $\mathcal{F}$.

(1) Since each $\mathcal{F}_i$ is a σ-field, we have $\emptyset \in \mathcal{F}_i$. Hence $\emptyset \in \cap_i \mathcal{F}_i$.

(2) If $A \in \mathcal{F} = \cap_i \mathcal{F}_i$ then $A \in \mathcal{F}_i$ for each i. Since each $\mathcal{F}_i$ is a σ-field, $\Omega \setminus A \in \mathcal{F}_i$ for each i. Hence $\Omega \setminus A \in \cap_i \mathcal{F}_i$.

(3) If $A_j \in \mathcal{F}$ for all j, then $A_j \in \mathcal{F}_i$ for all i and j. Since each $\mathcal{F}_i$ is a σ-field, $\cup_j A_j \in \mathcal{F}_i$ for all i. Hence $\cup_j A_j \in \cap_i \mathcal{F}_i$. ■

Corollary 2.1.6 *In particular, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ-fields, so is $\mathcal{F}_1 \cap \mathcal{F}_2$.*

Now, suppose that we have our Ω and we have a finite or countable collection of $E_1, E_2, \ldots \subseteq \Omega$, which we want to be events. Let $\mathscr{F}$ be the set of all σ-fields that contain $E_1, E_2, \ldots$. We enumerate $\mathscr{F}$ as $\mathscr{F} = \{\mathcal{F}_i\,;\, i \in I\}$, and apply Lemma 2.1.5. We thus obtain a σ-field $\mathcal{F}$, which contains all the events that we wanted.

The key point here is that $\mathcal{F}$ is the smallest σ-field that has $E_1, E_2, \ldots$ as events. To see why, note that by (2.1), $\mathcal{F}$ is contained inside any σ-field $\mathcal{F}'$ which has $E_1, E_2, \ldots$ as events.

Definition 2.1.7 *Let $E_1, E_2, \ldots$ be subsets of Ω. We write $\sigma(E_1, E_2, \ldots,)$ for the smallest σ-field containing $E_1, E_2, \ldots$.*

With Ω as any set, and $A \subseteq \Omega$, our example $\{\emptyset, A, \Omega \setminus A, \Omega\}$ is clearly $\sigma(A)$. In general, though, the point of Definition 2.1.7 is that we know useful σ-fields exist without having to construct them explicitly.

In the same style, if $\mathcal{F}_1, \mathcal{F}_2 \ldots$ are σ-fields then we write $\sigma(\mathcal{F}_1, \mathcal{F}_2, \ldots)$ for the smallest σ-algebra with respect to which all events in $\mathcal{F}_1, \mathcal{F}_2, \ldots$ are measurable.

From Definition 2.1.1 and 2.1.3 we can deduce all the 'usual' properties of probability. For example:

- If $A \in \mathcal{F}$ then $\Omega \setminus A \in \mathcal{F}$, and since $\Omega = A \cup (\Omega \setminus A)$ we have $= 1 = \mathbb{P}[A] + \mathbb{P}[\Omega \setminus A]$.
- If $A, B \in \mathcal{F}$ and $A \subseteq B$ then we can write $B = A \cup (B \setminus A)$, which gives us that $\mathbb{P}[B] = \mathbb{P}[B \setminus A] + P[A]$, which implies that $\mathbb{P}[A] \leq \mathbb{P}[B]$.

And so on. In this course we are concerned with applying probability theory rather than with relating its properties right back to the definition of a probability space; but you should realize that it is always possible to do so.

Definitions 2.1.1 and 2.1.3 both involve countable unions. Its convenient to be able to use countable intersections too, for which we need the following lemma.

Lemma 2.1.8 *Let $A_1, A_2, \ldots \in \mathcal{F}$, where $\mathcal{F}$ is a σ-field. Then $\bigcap_{i=1}^\infty A_i \in \mathcal{F}$.*

Proof: We can write

$$\bigcap_{i=1}^\infty A_i = \bigcap_{i=1}^\infty \Omega \setminus (\Omega \setminus A_i) = \Omega \setminus \left(\bigcup_{i=1}^\infty \Omega \setminus A_i \right).$$

Since $\mathcal{F}$ is a σ-field, $\Omega \setminus A_i \in \mathcal{F}$ for all i. Hence also $\bigcup_{i=1}^\infty \Omega \setminus A_i \in \mathcal{F}$, which in turn means that $\Omega \setminus (\bigcup_{i=1}^\infty \Omega \setminus A_i) \in \mathcal{F}$. ■

In general, *uncountable* unions and intersections of measurable sets need not be measurable. The reasons why we only allow countable unions/intersections in probability are complicated and beyond the scope of this course. Loosely speaking, the bigger we make $\mathcal{F}$, the harder it is to make a probability measure $\mathbb{P}$, because we need to define $\mathbb{P}[A]$ for all $A \in \mathcal{F}$ in a way that satisfies Definition 2.1.3. Allowing uncountable set operations would (in natural situations) result in $\mathcal{F}$ being so large that it would be *impossible* to find a suitable $\mathbb{P}$.

From now on, the symbols Ω, $\mathcal{F}$ and $\mathbb{P}$ always denote the three elements of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2.2 Random variables

Our probability space gives us a label $\omega \in \Omega$ for every possible outcome. Sometimes it is more convenient to think about a property of ω, rather than about ω itself. For this, we use a random variable, $X : \Omega \to \mathbb{R}$. For each outcome $\omega \in \Omega$, the value of $X(\omega)$ is a property of the outcome.

For example, let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$. We might be interested in the property

$$X(\omega) = \begin{cases} 0 & \text{if } X \text{ is odd,} \\ 1 & \text{if } X \text{ is even,} \end{cases}$$

We write

$$X^{-1}(A) = \{\omega \in \Omega\,;\, X(\omega) \in A\},$$

for $A \subseteq \mathbb{R}$, which is called the *pre-image* of A under X. In words, $X^{-1}(A)$ is the set of outcomes ω for which the property $X(\omega)$ falls inside the set A. In our example above $X^{-1}(\{0\}) = \{1, 3, 5\}$, $X^{-1}(\{1\}) = \{2, 4, 6\}$ and $X^{-1}(\{0, 1\}) = \{1, 2, 3, 4, 5, 6\}$.

It is common to write $X^{-1}(a)$ in place of $X^{-1}(\{a\})$, because it makes easier reading. Similarly, for an interval $(a, b) \subseteq \mathbb{R}$ we write $X^{-1}(a, b)$ in place of $X^{-1}\big((a, b)\big)$.

Definition 2.2.1 *Let $\mathcal{G}$ be a σ-field. A function $X : \Omega \to \mathbb{R}$ is said to be $\mathcal{G}$-measurable if*

$$\textit{for all subintervals } I \subseteq \mathbb{R}, \quad \textit{we have } X^{-1}(I) \in \mathcal{G}.$$

If it is already clear which σ-field $\mathcal{G}$ should be used in the definition, which simply say that X is *measurable*. We will often shorten this to writing simply $X \in m\mathcal{G}$. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that $X : \Omega \to \mathbb{R}$ is a *random variable* is X if $\mathcal{F}$-measurable.

The relationship to the notation you usually used in probability is that $\mathbb{P}[X \in A]$ means $\mathbb{P}[X^{-1}(A)]$, so as e.g.

$$\mathbb{P}\left[a < X < b\right] = \mathbb{P}[\omega \in \Omega\,;\, X(\omega) \in (a, b)] = \mathbb{P}\left[X^{-1}(a, b)\right].$$

Similarly, for any $a \in \mathbb{R}$, the set $\{a\} = [a, a]$ which is a subinterval, so

$$\mathbb{P}[X = a] = \mathbb{P}[\omega \in \Omega\,;\, X(\omega) = a] = \mathbb{P}[X^{-1}(a)].$$

We tend to prefer writing $\mathbb{P}[X = a]$ instead of $\mathbb{P}[X^{-1}(a)]$ because we like to think of X as an object that takes a random value, so $\mathbb{P}[X = a]$ is more intuitive.

The key point in Definition 2.2.1 is that, *when we choose how big we want our $\mathcal{F}$ to be, we are also choosing which functions $X : \Omega \to \mathbb{R}$ are random variables.* This will become very important to us later in the course.

For example, suppose we toss a coin twice, with $\Omega = \{HH, HT, TH, TT\}$ as in Example 2.1.2. If we chose $\mathcal{F} = \mathcal{P}(\Omega)$ then any subset of Ω is $\mathcal{F}$-measurable, and consequently any function $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable. However, suppose we choose instead

$$\mathcal{G} = \big\{\Omega, \Omega \setminus \{HH\}, \{HH\}, \emptyset\big\}$$

(as we did in Example 2.1.2). Then if we look at function

$$X(\omega) = \text{ the total number of tails which occurred}$$

we have

$$X^{-1}[0,1] = \{HH, HT, TH\} \notin \mathcal{G}.$$

So X is not $\mathcal{G}$-measurable. However, the function

$$Y(\omega) = \begin{cases} 0 & \text{if both coins were heads} \\ 1 & \text{otherwise} \end{cases}$$

is $\mathcal{G}$-measurable; to see this we can list

$$Y^{-1}(I) = \begin{cases} \emptyset & \text{if } 0,1 \notin I \\ \{HH\} & \text{if } 0 \in I \text{ and } 1 \notin I \\ \Omega \setminus \{HH\} & \text{if } 0 \notin I \text{ and } 1 \in I \\ \Omega & \text{if } 0,1 \in I. \end{cases} \tag{2.2}$$

The interaction between random variables and σ-fields can be summarised as follows:

$$\boxed{\sigma\text{-field } \mathcal{F}} \quad \leftrightarrow \quad \boxed{\text{which information we care about}}$$
$$\boxed{X \text{ is } \mathcal{F}\text{-measurable}} \quad \leftrightarrow \quad \boxed{X \text{ depends only on information that we care about}}$$

Rigorously, if we want to check that X is $\mathcal{F}$-measurable, we have to check that $X^{-1}(I) \in \mathcal{F}$ for every subinterval of $I \subseteq \mathbb{R}$. This can be tedious[1]. Fortunately, we will shortly see that, in practice, there is rarely any need to do so. What is important for us is to understand the role played by a σ-field.

[1]There are measure theoretic tools to make the job easier, but they are beyond the scope of our course.

2.2.1 σ-fields generated by random variables

We can think of random variables as containing information, because their values tell us something about the result of the experiment. We can express this idea formally: there is a natural σ-field associated to each function $X : \Omega \to \mathbb{R}$.

Definition 2.2.2 *The σ-field generated by X, denoted $\sigma(X)$, is*

$$\sigma(X) = \sigma(X^{-1}(I)\,;\, I \textit{ is a subinterval of } \mathbb{R}).$$

In words, $\sigma(X)$ is the σ-field generated by the sets $X^{-1}(I)$ for intervals I. The intuition is that $\sigma(X)$ is the smallest σ-field of events on which the random behaviour of X depends.

For example, consider throwing a fair die. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, let $\mathcal{F} = \mathcal{P}(\Omega)$. Let

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is odd} \\ 2 & \text{if } \omega \text{ is even.} \end{cases}$$

Then $X(\omega) \in \{1, 2\}$, with pre-images $X^{-1}(1) = \{1, 3, 5\}$ and $X^{-1}(2) = \{2, 4, 6\}$. The smallest σ-field that contains both of these subsets is

$$\sigma(X) = \Big\{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\Big\}.$$

In general, if X takes lots of different values, $\sigma(X)$ could be very big and we would have no hope of writing it out explicitly. Here's another example: suppose that

$$Y(\omega) = \begin{cases} 1 & \text{if } \omega = 1, \\ 2 & \text{if } \omega = 2, \\ 3 & \text{if } \omega \geq 3. \end{cases}$$

Then $Y(\omega) \in \{1, 2, 3\}$ with pre-images $Y^{-1}(1) = \{1\}$, $Y^{-1}(2) = \{2\}$ and $Y^{-1}(3) = \{3, 4, 5, 6\}$. The smallest σ-field containing these three subsets is

$$\sigma(Y) = \Big\{\emptyset, \{1\}, \{2\}, \{3, 4, 5, 6\}, \{1, 2\}, \{1, 3, 4, 5, 6\}, \{2, 3, 4, 5, 6\}, \Omega\Big\}.$$

It's natural that X should be measurable with respect to the σ-field that contains precisely the information on which X depends. Formally:

Lemma 2.2.3 *Let $X : \Omega \to \mathbb{R}$. Then X is $\sigma(X)$-measurable.*

PROOF: Let I be a subinterval of $\mathbb{R}$. Then, by definition of $\sigma(X)$, we have that $X^{-1}(I) \in \sigma(X)$ for all I. ■

More generally, for a finite or countable set of random variables $X_1, X_2, \ldots$ we define $\sigma(X_1, X_2, \ldots)$ to be $\sigma(X_1^{-1}(I), X_2^{-1}(I), \ldots\,;\, I$ is a subinterval of $\mathbb{R})$. The intuition is the same: $\sigma(X_1, X_2, \ldots)$ corresponds to the information jointly contains in $X_1, X_2, \ldots$.

2.2.2 Combining random variables

Given a collection of random variables, it is useful to be able to construct other random variables from them. To do so we have the following proposition. Since we will eventually deal with more than one σ-field at once, it is useful to express this idea for a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$.

Proposition 2.2.4 *Let $\alpha \in \mathbb{R}$ and let $X, Y, X_1, X_2, \dots$ be $\mathcal{G}$-measurable functions from $\Omega \to \mathbb{R}$. Then*

$$\alpha, \quad \alpha X, \quad X + Y, \quad XY, \quad 1/X, \tag{2.3}$$

are all $\mathcal{G}$-measurable. Further, if X_∞ given by

$$X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$$

exists for all ω, then X_∞ is $\mathcal{G}$-measurable.

Essentially, every natural way of combining random variables together leads to other random variables. Proposition 2.2.4 can usually be used to show this.

For example, if X is a random variable then so is $\frac{X^2+X}{2}$. For a more difficult example, suppose that X is a random variable and let $Y = e^X$, which means that $Y(\omega) = \lim_{n\to\infty} \sum_{i=0}^n \frac{X(\omega)^i}{i!}$. Recall that we know from analysis that this limit exists since $e^x = \lim_{n\to\infty} \sum_{i=0}^n \frac{x^i}{i!}$ exists for all $x \in \mathbb{R}$. Each of the partial sums

$$Y_n(\omega) = \sum_{i=0}^n \frac{X(\omega)^i}{i!} = 1 + X + \frac{X^2}{2} + \ldots + \frac{X^n}{n!}$$

is a random variable (we could use (2.3) repeatedly to show this) and, since the limit exists, $Y(\omega) = \lim_{n\to\infty} Y_n(\omega)$ is measurable.

In general, if X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is any 'sensible' function then $g(X)$ is also a random variable. This includes polynomials, powers, all trig functions, all monotone functions, all piecewise linear functions, all integrals/derivatives, etc etc.

2.2.3 Independence

We can express the concept of independence, which you already know about for random variables, in terms of σ-fields. Recall that two events $E_1, E_2 \in \mathcal{F}$ are said to be independent if

$$\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[E_1]\mathbb{P}[E_2].$$

Using σ-fields, we have a consistent way of defining independence, for both random variables and events.

Definition 2.2.5 *Sub-σ-fields $\mathcal{G}_1, \mathcal{G}_2$ of $\mathcal{F}$ are said to be* independent *if, whenever $G_i \in \mathcal{G}_i$, $i = 1, 2$, we have $\mathbb{P}(G_1 \cap G_2) = \mathbb{P}(G_1)\mathbb{P}(G_2)$.*

Random variables X_1 and X_2 are said to be independent if the σ-fields $\sigma(X_1)$ and $\sigma(X_2)$ are independent.

Events E_1 and E_2 are said to be independent if $\sigma(E_1)$ and $\sigma(E_2)$ are independent.

($\star$) It can be checked that, for events and random variables, this definition is equivalent to the definitions you may have seen in earlier courses.

2.3 Two kinds of examples

In this section we consolidate our knowledge from the previous two sections by looking at two important contrasting examples.

2.3.1 Finite Ω

Let $n \in \mathbb{N}$, and let $\Omega = \{x_1, x_2, \dots x_n\}$ be a finite set. Let $\mathcal{F} = \mathcal{P}(\Omega)$, which is also a finite set. We have seen how it is possible to construct other σ-fields on Ω too. Since $\mathcal{F}$ contains every subset of Ω, any σ-field on Ω is a sub-σ-field of $\mathcal{F}$.

In this case we can define a probability measure on Ω by choosing a finite sequence $a_1, a_2, \dots, a_n$ such that each $a_i \in [0,1]$ and $\sum_1^n a_i = 1$. We set $\mathbb{P}[x_i] = a_i$. This naturally extends to defining $\mathbb{P}[A]$ for any subset $A \subseteq \Omega$, by setting

$$\mathbb{P}[A] = \sum_{\{i\,;\,x_i \in A\}} \mathbb{P}[x_i] = \sum_{\{i\,;\,x_i \in A\}} a_i. \tag{2.4}$$

It is hopefully obvious (and tedious to check) that, with this definition, $\mathbb{P}$ is a probability measure. Consequently $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

All experiments with only finitely many outcomes fit into this category of examples. We have already seen several of them.

- Roll a biased die. Choose $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and define $\mathbb{P}$ by setting $\mathbb{P}[i] = \frac{1}{8}$ for $i = 1,2,3,4,5$ and $\mathbb{P}[6] = \frac{3}{8}$.
- Toss a fair coin twice, independently. Choose $\Omega = \{HH, TH, HT, TT\}$, $\mathcal{F} = \mathcal{P}(\Omega)$. Define $\mathbb{P}$ by setting $\mathbb{P}[**] = \frac{1}{4}$, where each instance of $*$ denotes either H or T.

For a sub-σ-field $\mathcal{G}$ of $\mathcal{F}$, the triplet $(\Omega, \mathcal{G}, \mathbb{P}_{\mathcal{G}})$ is also a probability space. Here $\mathbb{P}_{\mathcal{G}} : \mathcal{G} \to [0,1]$ simply means $\mathbb{P}$ restricted to $\mathcal{G}$, i.e. $\mathbb{P}_{\mathcal{G}}[A] = \mathbb{P}[A]$.

If $\mathcal{G} \neq \mathcal{F}$, some random variables $X : \Omega \to \mathbb{R}$ are $\mathcal{G}$-measurable and others are not. Intuitively, a random variable X is $\mathcal{G}$-measurable if we can deduce the value of $X(\omega)$ from knowing only, for all $G \in \mathcal{G}$, if $\omega \in G$. Each $G \in \mathcal{G}$ represents a piece of information that $\mathcal{G}$ allows us access too (and this piece of information is whether or not $\omega \in \mathcal{G}$); if $\mathcal{G}$ gives us access to enough information then we can determine the value of $X(\omega)$ for all ω, in which case we say that X is $\mathcal{G}$-measurable.

Rigorously, to check if a given random variable is $\mathcal{G}$ measurable, we can either check the pre-images directly, or (usually better) use Proposition 2.2.4. To show that a given random variable X is not $\mathcal{G}$-measurable, we just need to find an interval $I \in \mathbb{R}$ such that $X^{-1}(I) \notin \mathcal{G}$.

2.3.2 An example with infinite Ω

Now we flex our muscles a bit, and look at an example where Ω is infinite. We toss a coin infinitely many times, then $\Omega = \{H,T\}^{\mathbb{N}}$, meaning that we write an outcome as a sequence $\omega = \omega_1, \omega_2, \ldots$ where $\omega_i \in \{H,T\}$. We define the random variables $X_n(\omega) = \omega_n$, so as X_n represents the result (H or T) of the n^{th} throw. We take

$$\mathcal{F} = \sigma(X_1, X_2, \ldots)$$

i.e. $\mathcal{F}$ is smallest σ-field with respect to which all the X_n are random variables. Then

$$\begin{aligned}
\sigma(X_1) &= \{\emptyset, \{H**\ast\ldots\}, \{T**\ast\ldots\}, \Omega\} \\
\sigma(X_1, X_2) &= \sigma\Big(\{HH**\ldots\}, \{TH**\ldots\}, \{HT**\ldots\}, \{TT**\ldots\}\Big) \\
&= \Big\{\emptyset, \{HH**\ldots\}, \{TH**\ldots\}, \{HT**\ldots\}, \{TT**\ldots\}, \\
&\qquad \{H**\ast\ldots\}, \{T**\ast\ldots\}, \{*H**\ldots\}, \{*T**\ldots\}, \left\{\begin{array}{c} HH**\ldots \\ TT**\ldots \end{array}\right\}, \left\{\begin{array}{c} HT**\ldots \\ TH**\ldots \end{array}\right\}, \\
&\qquad \{HH**\ldots\}^c, \{TH**\ldots\}^c, \{HT**\ldots\}^c, \{TT**\ldots\}^c, \Omega\Big\},
\end{aligned}$$

where $*$ means that it can take on either H or T, so $\{H**\ast\ldots\} = \{\omega : \omega_1 = H\}$.

With the information available to us in $\sigma(X_1, X_2)$, we can distinguish between ω's where the first or second outcomes are different. But if two ω's have the same first and second outcomes, they fall into exactly the same subset(s) of $\sigma(X_1, X_2)$. Consequently, if a random variable depends on anything more than the first and second outcomes, it will not be $\sigma(X_1, X_2)$ measurable.

It is not immediately clear if we can define a probability measure on $\mathcal{F}$! Since Ω is uncountable, we cannot follow the scheme in Section 2.3.1 and define $\mathbb{P}$ in terms of $\mathbb{P}[\omega]$ for each individual $\omega \in \Omega$. Equation (2.4) simply would not make sense; there is no such thing as an uncountable sum.

To define a probability measure in this case requires a significant amount of machinery from measure theory. It is outside of the scope of this course. For our purposes, whenever we need to use an infinite Ω you will be *given* a probability measure and some of its helpful properties. For example, in this case there exists a probability measure $\mathbb{P} : \mathcal{F} \to [0,1]$ such that

- $\mathbb{P}[X_n = H] = \mathbb{P}[X_n = T] = \frac{1}{2}$ for all $n \in \mathbb{N}$.
- Each X_n is independent.

From this, you can work with $\mathbb{P}$ without having to know how $\mathbb{P}$ was constructed. You don't even need to know exactly which subsets of Ω are in $\mathcal{F}$, because Proposition 2.2.4 gives you access to plenty of random variables.

Remark 2.3.1 *($\star$) In this case it turns out that $\mathcal{F}$ is much smaller than $\mathcal{P}(\Omega)$. In fact, if we tried to take $\mathcal{F} = \mathcal{P}(\Omega)$, we would (after some significant effort) discover that there is no probability measure $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ satisfying the two conditions we wanted above for $\mathbb{P}$. This is irritating, and we just have to live with it.*

2.3.3 Almost surely

In the example from Section 2.3.2 we used $\Omega = \{H, T\}^{\mathbb{N}}$, which is the set of all sequences made up of Hs and Ts. Our probability measure was independent, fair, coin tosses and we used the random variable X_n for the n^{th} toss.

Let's examine this example a bit. First let us note that, for any individual sequence $\omega_1, \omega_2, \ldots$ of heads and tails, by independence

$$\mathbb{P}[X_1 = \omega_1, X_2 = \omega_2, \ldots] = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \ldots = 0.$$

So every element of Ω has probability zero. This is not a problem – if we take enough elements of Ω together then we do get non-zero probabilities, for example

$$\mathbb{P}[X_1 = H] = \mathbb{P}\left[\omega \in \Omega \text{ such that } \omega_1 = H\right] = \frac{1}{2}$$

which is not surprising.

The probability that we never throw a head is

$$\mathbb{P}[\text{for all } n, X_n = T] = \frac{1}{2} \cdot \frac{1}{2} \ldots = 0$$

which means that the probability that we eventually throw a head is

$$\mathbb{P}[\text{for some } n, X_n = H] = 1 - \mathbb{P}[\text{for all } n, X_n = T] = 1.$$

So, the event $\{\text{for some } n, X_n = H\}$ has probability 1, but is *not equal to* the whole sample space Ω. To handle this situation we have a piece of terminology.

Definition 2.3.2 *If the event E has $\mathbb{P}[E] = 1$, then we say E occurs almost surely.*

So, we would say that 'almost surely, our coin will eventually throw a head'. We might say that '$Y \leq 1$' almost surely, to mean that $\mathbb{P}[Y \leq 1] = 1$. This piece of terminology will be used very frequently from now on. We might sometimes say that an event 'almost always' happens, with the same meaning.

For another example, suppose that we define q_n^H and q_n^T to be the proportion of heads and, respectively, tails in the random sequence $X_1, X_2, \ldots, X_n$. Formally, this means that

$$q_n^H = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i = H\} \quad \text{and} \quad q_n^T = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i = T\}.$$

Of course $q_n^H + q_n^T = 1$.

The random variables $\mathbb{1}\{X_i = H\}$ are i.i.d. with $\mathbb{E}[\mathbb{1}\{X_i = H\}] = \frac{1}{2}$, hence by Theorem 1.1.1 we have $\mathbb{P}[q_n^H \to \frac{1}{2} \text{ as } n \to \infty] = 1$, and by the same argument we have also $\mathbb{P}[q_n^T \to \frac{1}{2} \text{ as } n \to \infty] = 1$. In words, this means that in the long run half our tosses will be tails and half will be heads (which makes sense - our coin is fair). We say that the event

$$E = \left\{ \lim_{n\to\infty} q_n^H = \frac{1}{2} \text{ and } \lim_{n\to\infty} q_n^T = \frac{1}{2} \right\}$$

occurs almost surely.

There are many many examples of sequences (e.g. $HHTHHTHHT\ldots$) that don't have $q_n^T \to \frac{1}{2}$ and $q_n^H \to \frac{1}{2}$. We might think of the set E as being only a 'small' subset of Ω, but it has probability one.

2.4 Expectation

There is only one part of the 'usual machinery' for probability that we haven't yet discussed, namely expectation.

Recall that the expectation of a discrete random variable X that takes the values $\{x_i : i \in \mathbb{N}\}$ is given by

$$\mathbb{E}[X] = \sum_{x_i} x_i \mathbb{P}[X = x_i]. \tag{2.5}$$

For a continuous random variables, the expectation uses an integral against the probability density function,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx. \tag{2.6}$$

Recall also that it is possible for limits (i.e. infinite sums) and integrals to be infinite, or not exist at all.

We are now conscious of the general definition of a random variable X, as an $\mathcal{F}$-measurable function from Ω to $\mathbb{R}$. There are many random variables that are neither discrete nor continuous, and for such cases (2.5) and (2.6) are not valid; we need a more general approach.

With *Lebesgue integration*, the expectation $\mathbb{E}$ can be defined using a single definition that works for both discrete and continuous (and other more exotic) random variables. This definition relies heavily on analysis and is well beyond the scope of this course. Instead, Lebesgue integration is covered in MAS350/451/6051.

For purposes of this course, what you should know is: $\mathbb{E}[X]$ is defined for all X such that either

1. $X \geq 0$, in which case it is possible that $\mathbb{E}[X] = \infty$,
2. general X for which $\mathbb{E}[|X|] < \infty$.

The point here is that we are prepared to allow ourselves to write $\mathbb{E}[X] = \infty$ (e.g. when the sum or integral in (2.5) or (2.6) tends to ∞) *provided* that $X \geq 0$. We are not prepared to allow expectations to equal $-\infty$, because we have to avoid nonsensical '$\infty - \infty$' situations.

You may still use (2.5) and (2.6), in the discrete/continuous cases. You may also assume that all the 'standard' properties of $\mathbb{E}$ hold:

Proposition 2.4.1 *For random variables X, Y:*

(Linearity) *If $a, b \in \mathbb{R}$ then $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.*
(Independence) *If X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.*
(Absolute values) *$|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.*
(Monotonicity) *If $X \leq Y$ then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*
(Positivity) *If $X \geq 0$ and $\mathbb{E}[X] = 0$ then $\mathbb{P}[X = 0] = 1$.*

You should become familiar with any of the properties that you are not already used to using. The proofs of the first four properties are part of the formal construction of $\mathbb{E}$ and are not part of our course. Proving the last property is one of the challenge questions, see **2.9**.

2.4.1 Indicator functions

One important type of random variable is an indicator function. Let $A \in \mathcal{F}$, then the indicator function of A is the function

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

The indicator function is used to tell if an event occurred (in which case it is 1) or did not occur (in which case it is 0). It is useful to remember that

$$\mathbb{P}[A] = \mathbb{E}[\mathbb{1}_A].$$

We will sometimes not put the A as a subscript and write e.g. $\mathbb{1}\{X < 0\}$ for the indicator function of the event that $X < 0$.

As usual, let $\mathcal{G}$ denote a sub σ-field of $\mathcal{F}$.

Lemma 2.4.2 *Let $A \in \mathcal{G}$. Then the function $\mathbb{1}_A$ is $\mathcal{G}$-measurable.*

Proof: Let us write $Y = \mathbb{1}_A$. For any subinterval $I \subseteq \mathbb{R}$,

$$Y^{-1}(I) = \begin{cases} \emptyset & \text{if } 0, 1 \notin I \\ A & \text{if } 0 \notin I \text{ and } 1 \in I \\ \Omega \setminus A & \text{if } 0 \in I \text{ and } 1 \notin I \\ \Omega & \text{if } 0, 1 \in I. \end{cases}$$

In all cases we have $Y^{-1}(I) \in \mathcal{F}$. ■

Remark 2.4.3 ($\star$) *More generally, suppose that $X : \Omega \to \mathbb{R}$ is a function that takes only finite many values, say $X(\omega) \in \{a_1, a_2, \ldots, a_n\}$. Then X is $\mathcal{F}$-measurable if and only if $X^{-1}(a_i) \in \mathcal{F}$ for all i. The proof is similar to Lemma 2.4.2.*

Indicator functions allow to *condition*, meaning that we can break up a random variable into two cases. For example, if $a \in \mathbb{R}$ we might write

$$X = X\mathbb{1}_{\{X<a\}} + X\mathbb{1}_{\{X\geq a\}}. \tag{2.7}$$

On the right hand side, precisely one of the two terms is non-zero. If the first term is non-zero then we can assume $X < a$, if the second is non-zero then we can assume $X \geq a$. This is very useful, for example:

Lemma 2.4.4 (Markov's Inequality) *Let $a > 0$ and let X be a random variable such that $X \geq 0$. Then*

$$\mathbb{P}[X \geq a] \leq \frac{1}{a}\mathbb{E}[X].$$

Proof: From (2.7) we have

$$X \geq X\mathbb{1}_{\{X\geq a\}} \geq a\mathbb{1}\{X \geq a\}.$$

Hence, using monotonicity of $\mathbb{E}$, we have

$$\mathbb{E}[X] \geq \mathbb{E}[a\mathbb{1}_{\{X\geq a\}}] = a\mathbb{E}[1_{\{X\geq a\}}] = a\mathbb{P}[X \geq a].$$

Dividing through by a finishes the proof. ■

2.4.2 L^p spaces

It will often be important to us to check whether a random variable X has finite mean and variance. Some random variables do not, see exercise **2.6** (or MAS223) for example. Random variables with finite mean and variances are easier to work with than those which don't, and many of the results in this course require these conditions.

We need some notation:

Definition 2.4.5 *Let $p \in [1, \infty)$. We say say that $X \in L^p$ if $\mathbb{E}[|X|^p] < \infty$.*

In this course, we will only be interested in the cases $p = 1$ and $p = 2$. In order to understand a little about these two spaces, let us prove an inequality:

$$\begin{aligned}
\mathbb{E}[|X|] &= \mathbb{E}[|X|\mathbb{1}_{\{|X|<1\}}] + \mathbb{E}[|X|\mathbb{1}_{\{|X|\geq 1\}}] \\
&\leq 1 + \mathbb{E}[X^2\mathbb{1}_{\{X\geq 1\}}] \\
&\leq 1 + \mathbb{E}[X^2]. \qquad (2.8)
\end{aligned}$$

Here, in the first line we condition using the indicator function. To deduce the second line, the key point is $x^2 \geq x$ only if $x \geq 1$; for the first term we note that if $|X|\mathbb{1}_{\{|X|<1\}} \leq 1$ and use monotonicity of $\mathbb{E}$, for the second term we use that $|X|\mathbb{1}_{\{|X|\geq 1\}} \leq X^2\mathbb{1}_{\{|X|\geq 1\}}$ and again use monotonicity of $\mathbb{E}$. For the last line we use that $X^2\mathbb{1}_{\{|X|\geq 1\}} \leq X^2$.

Coming back to L^p spaces, we can now state the following set of useful properties:

1. By definition, L^1 is the set of random variables for which $\mathbb{E}[|X|]$ is finite.
2. From (2.8), if $X \in L^2$ then also $X \in L^1$.
3. L^2 is the set of random variables with finite variance. To show this fact, we use that $\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, so $\text{var}(X) < \infty \Leftrightarrow \mathbb{E}[X^2] < \infty$.

Often, to check if $X \in L^p$ we must calculate $\mathbb{E}[|X|^p]$. A special case where it is automatic is the following.

Definition 2.4.6 *We say that a random variable X is bounded if there exists (deterministic) $c \in \mathbb{R}$ such that $|X| \leq c$.*

If X is bounded, then using monotonicity we have $\mathbb{E}[|X|^p] \leq \mathbb{E}[c^p] = c^p < \infty$, which means that $X \in L^p$, for all p.

2.5 Exercises

On probability spaces

2.1 Consider the experiment of throwing two dice, then recording the uppermost faces of both dice. Write down a suitable sample space Ω and suggest an appropriate σ-field $\mathcal{F}$.

2.2 Let $\Omega = \{1, 2, 3\}$. Let

$$\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\},$$
$$\mathcal{F}' = \{\emptyset, \{2\}, \{1, 3\}, \{1, 2, 3\}\}.$$

(a) Show that $\mathcal{F}$ and $\mathcal{F}'$ are both σ-fields.

(b) Show that $\mathcal{F} \cup \mathcal{F}'$ is not a σ-field, but that $\mathcal{F} \cap \mathcal{F}$ is a σ-field.

(c) Let $X : \Omega \to \mathbb{R}$ be defined by

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = 1 \\ 2 & \text{if } \omega = 2 \\ 1 & \text{if } \omega = 3 \end{cases}$$

Is X measurable with respect to $\mathcal{F}$? Is X measurable with respect to $\mathcal{F}'$?

2.3 Let $\Omega = \{H, T\}^{\mathbb{N}}$ be the probability space from Section 2.3.2, corresponding to an infinite sequence of independent fair coin tosses $(X_n)_{n=1}^{\infty}$.

(a) Fix $m \in \mathbb{N}$. Show that the probability that that random sequence $X_1, X_2, \ldots,$ contains precisely m heads is zero.

(b) Deduce that, almost surely, the sequence $X_1, X_2, \ldots$ contains infinitely many heads and infinitely many tails.

On random variables

2.4 Let $\Omega = \{HH, HT, TH, TT\}$, representing two coin tosses. Define X to be the total number of heads shown. Write down all the events in $\sigma(X)$.

2.5 Let X be a random variable. Explain why $\frac{X}{X^2+1}$ and $\sin(X)$ are also random variables.

2.6 Let X be a random variable with the probability density function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 2x^{-3} & \text{if } x \in [1, \infty), \\ 0 & \text{otherwise.} \end{cases}$$

Show that $X \in L^1$ but $X \notin L^2$.

2.7 Let $1 \le p \le q < \infty$ and let $X \in L^q$. Show that $X \in L^p$.

2.8 Let $c \in \mathbb{R}$ and let X be a random variable. Show that $\sigma(X) = \sigma(c + X)$ and, if $c \neq 0$, that $\sigma(X) = \sigma(cX)$.

Challenge questions

2.9 Show that if $\mathbb{P}[X \ge 0] = 1$ and $\mathbb{E}[X] = 0$ then $\mathbb{P}[X = 0] = 1$.

Chapter 3

Conditional expectation and martingales

We will introduce conditional expectation, which provides us with a way to estimate random quantities based on only partial information. We will also introduce martingales, which are the mathematical way to capture the concept of a fair game.

3.1 Conditional expectation

Suppose X and Z are random variables that take on only finitely many values $\{x_1, \ldots, x_m\}$ and $\{z_1, \ldots, z_n\}$, respectively. In earlier courses, 'conditional expectation' was defined as follows:

$$\begin{aligned} \mathbb{P}[X = x_i \,|\, Z = z_j] &= \mathbb{P}[X = x_i, Z = z_j]/\mathbb{P}[Z = z_j] \\ \mathbb{E}[X \,|\, Z = z_j] &= \sum_i x_i \mathbb{P}[X = x_i \,|\, Z = z_j] \\ Y = \mathbb{E}[X \,|\, Z] \text{ where:} &\quad \text{if } Z(\omega) = z_j, \text{ then } Y(\omega) = \mathbb{E}[X \,|\, Z = z_j] \end{aligned} \tag{3.1}$$

You might also have seen a second definition, using probability density functions, for continuous random variables. These definitions are problematic, for several reasons, chiefly (1) its not immediately clear how the two definitions interact and (2) we don't want to be restricted to handling only discrete or only continuous random variables.

In this section, we define the conditional expectation of random variables using σ-fields. In this setting we are able to give a unified definition which is valid for general random variables. The definition is originally due to Kolmogorov (in 1933), and is sometimes referred to as Kolmogorov's conditional expectation. It is one of the most important concepts in modern probability theory.

Conditional expectation is a mathematical tool with the following function. We have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. However, $\mathcal{F}$ is large and we want to work with a sub-σ-algebra $\mathcal{G}$, instead. As a result, we want to have a random variable Y such that

1. Y is $\mathcal{G}$-measurable
2. Y is 'the best' way to approximate X with a $\mathcal{G}$-measurable random variable

The second statement on this wish-list does not fully make sense; there are many different ways in which we could compare X to a potential Y.

Why might we want to do this? Imagine we are conducting an experiment in which we gradually gain information about the result X. This corresponds to gradually seeing a larger and larger $\mathcal{G}$, with access to more and more information. At all times we want to keep a prediction of what the future looks like, based on the currently available information. This prediction is Y.

It turns out there is *only one* natural way in which to realize our wish-list (which is convenient, and somewhat surprising). It is the following:

Theorem 3.1.1 (Conditional Expectation) *Let X be an L^1 random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Then there exists a random variable $Y \in L^1$ such that*

1. *Y is $\mathcal{G}$-measurable,*

2. *for every $G \in \mathcal{G}$, we have $\mathbb{E}[Y \mathbb{1}_G] = \mathbb{E}[X \mathbb{1}_G]$.*

Moreover, if $Y' \in L^1$ is a second random variable satisfying these conditions, $\mathbb{P}[Y = Y'] = 1$.

The first and second statements here correspond respectively to the items on our wish-list.

Definition 3.1.2 *We refer to Y as (a version of) the conditional expectation of X given $\mathcal{G}$. and we write*

$$Y = \mathbb{E}[X \mid \mathcal{G}].$$

Since any two such Y are almost surely equal so we sometimes refer to Y simply as *the* conditional expectation of X. This is a slight abuse of notation, but it is commonplace and harmless.

Proof of Theorem 3.1.1 is beyond the scope of this course. Loosely speaking, there is an abstract recipe which constructs $\mathbb{E}[X|\mathcal{G}]$. It begins with the random variable X, and then averages out over all the information that is not accessible to $\mathcal{G}$, leaving only as much randomness as $\mathcal{G}$ can support, resulting in $\mathbb{E}[X|\mathcal{G}]$. In this sense the map $X \mapsto \mathbb{E}[X|\mathcal{G}]$ simplifies (i.e. reduces the amount of randomness in) X in a very particular way, to make it $\mathcal{G}$ measurable.

It is important to remember that $\mathbb{E}[X|\mathcal{G}]$ is (in general) a random variable. It is also important to remember that the two objects

$$\mathbb{E}[X|\mathcal{G}] \quad \text{and} \quad \mathbb{E}[X|Z = z]$$

are quite different. They are both useful. We will explore the connection between them in Section 3.1.1. Before doing so, let us look at a basic example.

Let X_1, X_2 be independent random variables such that $\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = 1] = \frac{1}{2}$. Set $\mathcal{F} = \sigma(X_1, X_2)$. We will show that

$$\mathbb{E}[X_1 + X_2 | \sigma(X_1)] = X_1. \tag{3.2}$$

To do so, we should check that X_1 satisfies the two conditions in Theorem 3.1.1, with

$$\begin{aligned} X &= X_1 + X_2 \\ Y &= X_1 \end{aligned}$$

$$\mathcal{G} = \sigma(X_1).$$

The first condition is immediate, since by Lemma 2.2.3 X_1 is $\sigma(X_1)$-measurable i.e. $X \in m\mathcal{G}$. To see the second condition, let $G \in \sigma(X_1)$. Then $\mathbb{1}_G \in \sigma(X_1)$ and $X_2 \in \sigma(X_2)$, which are independent, so 1_G and X_2 are independent. Hence

$$\begin{aligned}
\mathbb{E}[(X_1 + X_2)\mathbb{1}_G] &= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{E}[1_G X_2] \\
&= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{E}[1_G]\mathbb{E}[X_2] \\
&= \mathbb{E}[X_1\mathbb{1}_G] + \mathbb{P}[G].0 \\
&= \mathbb{E}[X_1\mathbb{1}_G].
\end{aligned}$$

This equation says precisely that $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[Y\mathbb{1}_G]$. We have now checked both conditions, so by Theorem 3.1.1 we have $\mathbb{E}[X|\mathcal{G}] = Y$, meaning that $\mathbb{E}[X_1 + X_2|\sigma(X_1)] = X_1$, which proves our claim in (3.2).

The intuition for this, which is plainly visible in our calculation, is that X_2 is independent of $\sigma(X_1)$ so, thinking of conditional expectation as an operation which averages out all randomness in $X = X_1 + X_2$ that is not $\mathcal{G} = \sigma(X_1)$ measurable, we would average out X_2 completely i.e. $\mathbb{E}[X_2] = 0$.

We could equally think of X_1 as being our best guess for $X_1 + X_2$, given only information in $\sigma(X_1)$, since $\mathbb{E}[X_2] = 0$. In general, guessing $\mathbb{E}[X|\mathcal{G}]$ is not so easy!

3.1.1 Relationship to the naive definition ($\star$)

Conditional expectation extends the 'naive' definition of (3.1). Naturally, the 'new' conditional expectation is much more general (and, moreover, it is what we require later in the course), but we should still take the time to relate it to the naive definition.

Remark 3.1.3 *This subsection is marked with a ($\star$), meaning that it is non-examinable. This is so as you can forget the old definition and remember the new one!*

To see the connection, we focus on the case where X, Z are random variables with finite sets of values $\{x_1, \ldots, x_n\}$, $\{z_1, \ldots, z_m\}$. Let Y be the naive version of conditional expectation defined in (3.1). That is,

$$Y(\omega) = \sum_j \mathbb{1}_{\{Z(\omega)=z_j\}} \mathbb{E}[X|Z = z_j].$$

We can use Theorem 3.1.1 to check that, in fact, Y is a version of $\mathbb{E}[X|\sigma(Z)]$. We want to check that Y satisfies the two properties listed in Theorem 3.1.1.

- Since Z only takes finitely many values $\{z_1, \ldots, z_m\}$, from the above equation we have that Y only takes finitely many values. These values are $\{y_1, \ldots, y_m\}$ where $y_j = \mathbb{E}[X|Z = z_j]$. We note

  $$\begin{aligned} Y^{-1}(y_j) &= \{\omega \in \Omega\,;\, Y(\omega) = \mathbb{E}[X|Z = z_j]\} \\ &= \{\omega \in \Omega\,;\, Z(\omega) = z_j\} \\ &= Z^{-1}(z_j) \in \sigma(Z). \end{aligned}$$

  This is sufficient (although we will omit the details) to show that Y is $\sigma(Z)$-measurable.

- We can calculate

  $$\begin{aligned} \mathbb{E}[Y \mathbb{1}\{Z = z_j\}] &= y_j \mathbb{E}[\mathbb{1}\{Z = z_j\}] \\ &= y_j \mathbb{P}[Z = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i | Z = z_j] \mathbb{P}[Z_j = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i \text{ and } Z = z_j] \\ &= \sum_{i,j} x_i \mathbb{1}_{\{Z=z_j\}} \mathbb{P}[X = x_i \text{ and } Z = z_j] \\ &= \mathbb{E}[X \mathbb{1}_{\{Z=z_j\}}]. \end{aligned}$$

  Properly, to check that Y satisfies the second property in Theorem 3.1.1, we need to check $\mathbb{E}[Y \mathbb{1}_G] = \mathbb{E}[X \mathbb{1}_G]$ for a general $G \in \sigma(Z)$ and not just $G = \{Z = z_j\}$. However, for reasons beyond the scope of this course, in this case (thanks to the fact that Z is finite) its enough to consider only G of the form $\{Z = z_j\}$.

Therefore, we have $Y = \mathbb{E}[X|\sigma(Z)]$ almost surely. In this course we favour writing $\mathbb{E}[X|\sigma(Z)]$ instead of $\mathbb{E}[X|Z]$, to make it clear that we are looking at conditional expectation with respect to a σ-field.

3.2 Properties of conditional expectation

In all but the easiest cases, calculating conditional expectations explicitly from Theorem 3.1.1 is not feasible. Instead, we are able to work with them via a set of useful properties, provided by the following proposition.

Proposition 3.2.1 *Let $\mathcal{G}, \mathcal{H}$ be sub-σ-fields of $\mathcal{F}$ and $X, Y, Z \in L^1$. Then, almost surely,*

(Linearity) $\mathbb{E}[a_1X_1 + a_2X_2 \,|\, \mathcal{G}] = a_1\mathbb{E}[X_1 \,|\, \mathcal{G}] + a_2\mathbb{E}[X_2 \,|\, \mathcal{G}]$.
(Absolute values) $|\mathbb{E}[X \,|\, \mathcal{G}]| \leq \mathbb{E}[|X| \,|\, \mathcal{G}]$.
(Montonicity) *If* $X \leq Y$*, then* $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.
(Positivity) *If* $X \geq 0$ *and* $\mathbb{E}[X|\mathcal{G}] = 0$ *then* $X = 0$.
(Constants) *If* $a \in \mathbb{R}$ *(deterministic) then* $\mathbb{E}[a \,|\, \mathcal{G}] = a$.

(Measurability) *If* X *is* $\mathcal{G}$*-measurable, then* $\mathbb{E}[X \,|\, \mathcal{G}] = X$.
(Independence) *If* X *is independent of* $\mathcal{G}$ *then* $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]$.

(Taking out what is known) *If* Z *is* $\mathcal{G}$ *measurable, then* $\mathbb{E}[ZX \,|\, \mathcal{G}] = Z\mathbb{E}[X \,|\, \mathcal{G}]$.
(Tower) *If* $\mathcal{H} \subset \mathcal{G}$ *then* $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathbb{E}[X \,|\, \mathcal{H}]$.
(Taking $\mathbb{E}$) *It holds that* $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$.
(No information) *It holds that* $\mathbb{E}[X \,|\, \{\emptyset, \Omega\}] = \mathbb{E}[X]$,

PROOF: We will prove just two of these properties for ourselves, to give a feel for how the arguments work.

(Measurability): We check the two conditions of Theorem 3.1.1 with $Y = X$. By definition, X is $\mathcal{G}$-measurable. Since $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$, we are done.

(No information): Again, we check the two conditions of Theorem 3.1.1, now with $Y = \mathbb{E}[X]$. Since $\mathbb{E}[X]$ is deterministic, for any interval I we have $\{\mathbb{E}[X] \in I\}$ is either $\emptyset$ or Ω, hence $\mathbb{E}[X]$ is measurable with respect to $\mathcal{G} = \{\emptyset, \Omega\}$. Since $\mathcal{G}$ only has two elements we can check:

$$\begin{aligned} \mathbb{E}[X\mathbb{1}_\emptyset] &= \mathbb{E}[0] = 0 = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_\emptyset] && \text{since } \mathbb{1}_\emptyset = 0, \\ \mathbb{E}[X\mathbb{1}_\Omega] &= \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_\Omega] && \text{since } \mathbb{1}_\Omega = 1. \end{aligned}$$

Hence for all $G \in \mathcal{G}$ we have $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_G$. ■

Remark 3.2.2 ($\star$) *Although we have not proved many of the properties in Proposition 3.2.1, they are intuitive properties for conditional expectation to have.*

For example, in the taking out what is known property, we can think of Z as already being simple enough to be $\mathcal{G}$ measurable, so we'd expect that taking conditional expectation with respect to $\mathcal{G}$ doesn't need to affect it.

In the tower property for $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$, we start with X, simplify it to be $\mathcal{G}$ measurable and simplify it to be $\mathcal{H}$ measurable. But since $\mathcal{H} \subseteq \mathcal{G}$, we might as well have just simplified X enough to be $\mathcal{H}$ measurable in a single step, which would be $\mathbb{E}[X|\mathcal{H}]$.

Etc. It is a useful exercise for you to try and think of 'intuitive' arguments for the other properties too, so as you can easily remember them.

Remark 3.2.3 ($\star$) *You can find a more extensive list of the (many) properties of conditional expectation, with proofs of them all, in the book 'Probability with Martingales' by David Williams.*

The conditional expectation $Y = \mathbb{E}[X \,|\, \mathcal{G}]$ is the 'best least-squares estimator' of X, based on the information available in $\mathcal{G}$. We can state this rigorously and use our toolkit from Proposition 3.2.1 prove it. This demonstrates another way in which Y is 'the best' $\mathcal{G}$-measurable approximation to X.

Lemma 3.2.4 *Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Let X be an $\mathcal{F}$-measurable random variable and let $Y = \mathbb{E}[X|\mathcal{G}]$. Suppose that Y' is a $\mathcal{G}$-measurable, random variable. Then*

$$\mathbb{E}[(X-Y)^2] \le \mathbb{E}[(X-Y')^2].$$

Proof: We note that

$$\begin{aligned}\mathbb{E}[(X-Y')^2] &= \mathbb{E}[(X-Y+Y-Y')^2] \\ &= \mathbb{E}[(X-Y)^2] + 2\mathbb{E}[(X-Y)(Y-Y')] + \mathbb{E}[(Y-Y')^2]. \end{aligned} \tag{3.3}$$

In the middle term above, we can write

$$\begin{aligned}\mathbb{E}[(X-Y)(Y-Y')] &= \mathbb{E}[\mathbb{E}[(X-Y)(Y-Y')|\mathcal{G}]] \\ &= \mathbb{E}[(Y-Y')\mathbb{E}[X-Y|\mathcal{G}]] \\ &= \mathbb{E}[(Y-Y')(\mathbb{E}[X|\mathcal{G}]-Y)].\end{aligned}$$

Here, in the first step we used the tower property, in the second step we used Proposition 2.2.4 to tell us that $Y-Y'$ is $\mathcal{G}$-measurable, followed by the 'taking out what is known' rule. In the final step we used the linearity and measurability properties. Since $\mathbb{E}[X|\mathcal{G}] = Y$ almost surely, we obtain that $\mathbb{E}[(X-Y)(Y-Y')] = 0$. Hence, since $\mathbb{E}[(Y-Y')^2] \ge 0$, from (3.3) we obtain $\mathbb{E}[(X-Y')^2] \ge \mathbb{E}[(X-Y)^2]$. ∎

3.3 Martingales

In this section we introduce martingales, which are the mathematical representation of a 'fair game'. As usual, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

We refer to a sequence of random variables $(S_n)_{n=0}^{\infty}$ as a stochastic process. In this section of the course we only deal with discrete time stochastic processes. We say that a stochastic process (S_n) is *bounded* if there exists (deterministic) $c \in \mathbb{R}$ such that $|S_n(\omega)| \leq c$ for all n, ω.

We have previously discussed the idea of gradually learning more and more information about the outcome of some experiment, through seeing the information visible from gradually larger σ-fields. We formalize this concept as follows.

Definition 3.3.1 *A sequence of σ-fields $(\mathcal{F}_n)_{n=0}^{\infty}$ is known as a filtration if $\mathcal{F}_0 \subseteq \mathcal{F}_1 \ldots \subseteq \mathcal{F}$.*

Definition 3.3.2 *We say that a stochastic process $X = (X_n)$ is adapted to the filtration $(\mathcal{F}_n)$ if, for all n, X_n is $\mathcal{F}_n$ measurable.*

We should think of the filtration $\mathcal{F}_n$ as telling us which information we have access too at time $n = 1, 2, \ldots$. Thus, an adapted process is a process whose (random) value we know at all times $n \in \mathbb{N}$.

We are now ready to give the definition of a martingale.

Definition 3.3.3 *A process $M = (M_n)_{n=0}^{\infty}$ is a martingale if*

1. *if (M_n) is adapted,*
2. *$M_n \in L^1$ for all n,*
3. *$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = M_n$ almost surely, for all n.*

We say that M is a submartingale if, instead of 3, we have $\mathbb{E}[M_n \,|\, \mathcal{F}_{n-1}] \geq M_{n-1}$ almost surely. We say that M is a supermartingale if, instead of 3, we have $\mathbb{E}[M_n \,|\, \mathcal{F}_{n-1}] \leq M_{n-1}$ almost surely.

Remark 3.3.4 *The second condition in Definition 3.3.3 is needed for the third to make sense.*

Remark 3.3.5 *(M_n) is a martingale iff it is both a submartingale and a supermartingale.*

A martingale is the mathematical idealization of a fair game. It is best to understand what we mean by this through an example.

Let (X_n) be a sequence of i.i.d. random variables such that

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}.$$

Define $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Then $(\mathcal{F}_n)$ is a filtration. Define

$$S_n = \sum_{i=1}^{n} X_n$$

(and $S_0 = 0$). We can think of S_n as a game in the following way. At each time $n = 1, 2, \ldots$ we toss a coin. We win if the n^{th} round if the coin is heads, and lose if it is tails. Each time we win we score 1, each time we lose we score -1. Thus, S_n is our score after n rounds. The process S_n is often called a simple random walk.

We claim that S_n is a martingale. To see this, we check the three properties in the definition. (1) Since $X_1, X_2, \dots, X_n \in \sigma(X_1, \dots, X_n)$ we have that $S_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. (2) Since $|S_n| \leq n$ for all $n \in \mathbb{N}$, $\mathbb{E}[|S_n|] \leq n$ for all n, so $S_n \in L^1$ for all n. (3) We have

$$\begin{aligned}\mathbb{E}\left[S_{n+1} \,|\, \mathcal{F}_{n-1}\right] &= \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] + \mathbb{E}[S_n \,|\, \mathcal{F}_n] \\ &= \mathbb{E}[X_{n+1}] + S_n \\ &= S_n.\end{aligned}$$

Here, in the first line we used the linearity of conditional expectation. To deduce the second line we used the relationship between independence and conditional expectation (for the first term) and the tower rule (for the second term). To deduce the final line we used that $\mathbb{E}[X_{n+1}] = 0$.

At time n we have seen the result of rounds $1, 2, \dots, n$, so the information we currently have access to is given by $\mathcal{F}_n$. This means that at time n we know $S_1, \dots, S_n$. But we don't know S_{n+1}, because S_{n+1} is not $\mathcal{F}_n$-measurable. However, using our current information we can make our best guess at what S_{n+1} will be, which naturally is $\mathbb{E}[S_{n+1}|\mathcal{F}_n]$. Since the game is fair, in the future, on average we do not expect to win more than we lose, that is $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = S_n$.

In this course we will see many examples of martingales, and we will gradually build up an intuition for how to recognize a martingale. There is, however, one easy sufficient (but not necessary) condition under which we can recognize that a stochastic process is not a martingale.

Lemma 3.3.6 *Let $(\mathcal{F}_n)$ be a filtration and suppose that (M_n) is a martingale. Then*

$$\mathbb{E}[M_n] = \mathbb{E}[M_0]$$

for all $n \in \mathbb{N}$.

Proof: We have $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$. Taking expectations and using the 'taking $\mathbb{E}$' property from Proposition 3.2.1, we have $\mathbb{E}[M_{n+1}] = \mathbb{E}[M_n]$. The result follows by a trivial induction. ∎

Suppose, now, that (X_n) is an i.i.d. sequence of random variables such that $\mathbb{P}[X_i = 2] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Note that $\mathbb{E}[X_n] > 0$. Define S_n and $\mathcal{F}_n$ as before. Now, $\mathbb{E}[S_n] = \sum_1^n \mathbb{E}[X_n]$, which is not constant, so S_n is not a martingale. However, as before, S_n is $\mathcal{F}_n$-measurable, and $|S_n| \leq 2n$ so $S_n \in L^1$, essentially as before. We have

$$\begin{aligned}\mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}[S_n|\mathcal{F}_n] \\ &= \mathbb{E}[X_{n+1}] + S_n \\ &\geq S_n.\end{aligned}$$

Hence S_n is a submartingale.

In general, if (M_n) is a submartingale, then by definition $\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] \geq M_n$, so taking expectations gives us $\mathbb{E}[M_{n+1}] \geq \mathbb{E}[M_n]$. For supermartingales we get $\mathbb{E}[M_{n+1}] \leq \mathbb{E}[M_n]$. In words: submartingales, on average, increase, whereas supermartingales, on average, decrease. The use of super- and sub- is counter intuitive in this respect.

Remark 3.3.7 *Sometimes we will want to make it clear which filtration is being used in the definition of a martingale. To do so we might say that (M_n) is an $\mathcal{F}_n$-martingale', or that (M_n) is a martingale with respect to $\mathcal{F}_n$'. We use the same notation for super/sub-martingales.*

Our definition of a filtration and a martingale both make sense if we look at only a finite set of times $n = 1, \dots, N$. We sometimes also use the terms filtration and martingale in this situation.

We end this section with two more important general examples of martingales. You should check the conditions yourself, as exercise **3.4**.

Example 3.3.8 *Let (X_n) be a sequence of i.i.d. random variables such that $\mathbb{E}[X_n] = 1$ for all n, and there exists $c \in \mathbb{R}$ such that $|X_n| \leq c$ for all n. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then*

$$M_n = \prod_{i=1}^{n} X_n$$

is a martingale.

Example 3.3.9 *Let $Z \in L^1$ be a random variable and let $(\mathcal{F}_n)$ be a filtration. Then*

$$M_n = \mathbb{E}[Z|\mathcal{F}_n]$$

is a martingale.

3.4 Exercises

On conditional expectation

3.1 Let (X_n) be a sequence of independent identically distributed random variables, such that $\mathbb{P}[X_i = 1] = \frac{1}{2}$ and $\mathbb{P}[X_i = -1] = \frac{1}{2}$. Let

$$S_n = \sum_{i=1}^{n} X_i.$$

Find $\mathbb{E}[S_2 \,|\, \sigma(X_1)]$ and $\mathbb{E}[S_2^2 \,|\, \sigma(X_1)]$ in terms of X_1 and X_2.

3.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X, Y \in L^2$.

(a) Show that $X = Y$ almost surely if and only if $\mathbb{E}[(X - Y)^2] = 0$.

(b) Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Suppose that $\mathbb{E}[X \,|\, \mathcal{G}] = Y$ and $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$. Show that $X = Y$ almost surely.

On martingales

3.3 Let (X_n) be a sequence of independent random variables such that $\mathbb{P}[X_n = 2] = \frac{1}{3}$ and $\mathbb{P}[X_n = -1] = \frac{2}{3}$. Set $\mathcal{F}_n = \sigma(X_i\,;\, i \le n)$. Show that $S_n = \sum_{i=1}^{n} X_i$ is an $\mathcal{F}_n$ martingale.

3.4 Check that Examples 3.3.8 and 3.3.9, are martingales.

3.5 (a) Let (M_n) be an $\mathcal{F}_n$ martingale. Show that, for all $0 \le n \le m$, $\mathbb{E}[M_m \,|\, \mathcal{F}_n] = M_n$.

(b) Guess and state (without proof) the analogous result to (a) for submartingales.

3.6 Let (M_n) be a $\mathcal{F}_n$ martingale and suppose $M_n \in L^2$ for all n. Show that

$$\mathbb{E}[M_{n+1}^2 | \mathcal{F}_n] = M_n^2 + \mathbb{E}[(M_{n+1} - M_n)^2 | \mathcal{F}_n] \tag{3.4}$$

and deduce that (M_n^2) is a submartingale.

3.7 Let $X_0, X_1, \dots$ be a sequence of L^1 random variables. Let $\mathcal{F}_n$ be their generated filtration and suppose that $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = aX_n + bX_{n-1}$ for all $n \in \mathbb{N}$, where $a, b > 0$ and $a + b = 1$.

Find a value of $\alpha \in \mathbb{R}$ (in terms of a, b) for which $S_n = \alpha X_n + X_{n-1}$ is an $\mathcal{F}_n$ martingale.

Challenge questions

3.8 In the setting of **3.1**, show that $\mathbb{E}[X_1 \,|\, \sigma(S_n)] = \frac{S_n}{n}$.

Chapter 4

Stochastic processes

In this chapter we introduce stochastic processes, with a selection of examples that are commonly used as building blocks in stochastic modelling. We show that these stochastic processes are closely connected to martingales.

Definition 4.0.1 *A stochastic process (in discrete time) is a sequence $(X_n)_{n=0}^{\infty}$ of random variables. We think of n as 'time'.*

For example, a sequence of i.i.d. random variables is a stochastic process. A martingale is a stochastic process. A Markov chain (from MAS275, for those who took it) is a stochastic process. And so on.

For any stochastic process (X_n) the *natural* or *generated filtration* of (X_n) is the filtration given by

$$\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n).$$

Therefore, a random variable is $\mathcal{F}_m$ measurable if it depends only on the behaviour of our stochastic process up until time m.

From now on we adopt the convention (which is standard in the field of stochastic processes) that whenever we don't specify a filtration explicitly we mean to use the generated filtration.

4.1 Random walks

Random walks are stochastic processes that 'walk around' in space. We think of a particle that moves between vertices of $\mathbb{Z}$. At each step of time, the particle chooses at random to either move up or down, for example from x to $x+1$ or $x-1$.

4.1.1 Symmetric random walk

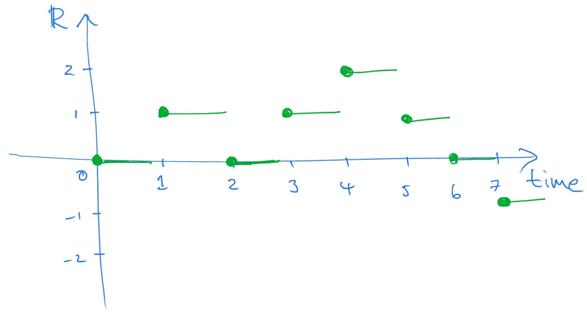
Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables where

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}. \tag{4.1}$$

The symmetric random walk is the stochastic process

$$S_n = \sum_{i=1}^{n} X_i.$$

By convention, this means that $S_0 = 0$. A *sample path* of S_n, which means a sample of the sequence $S_0, S_1, S_2, \ldots$, might look like:



Note that when time is discrete $t = 0, 1, 2, \ldots$ it is standard to draw the location of the random walk (and other stochastic processes) as constant in between integer time points.

Because of (4.1), the random walk is equally likely to move upwards or downwards. This case is known as the 'symmetric' random walk because, if $S_0 = 0$, the two stochastic processes S_n and $-S_n$ have the same distribution.

We have already seen (in Section 3.3) that S_n is a martingale, with respect to its generated filtration

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n) = \sigma(S_1, \ldots, S_n).$$

It should seem very natural that (S_n) is a martingale – going upwards as much as downwards is 'fair'.

4.1.2 Asymmetric random walk

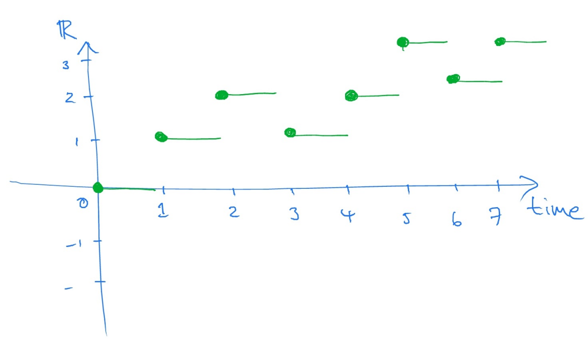
Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Let $p + q = 1$ with $p, q \in [0, 1]$, $p \neq q$ and suppose that

$$\mathbb{P}[X_i = 1] = p, \quad \mathbb{P}[X_i = -1] = q.$$

The asymmetric random walk is the stochastic process

$$S_n = \sum_{i=1}^{n} X_i.$$

The key difference to the symmetric random walk is that here we have $p \neq q$ (the symmetric random walk has $p = q = \frac{1}{2}$). The asymmetric random is more likely to step upwards than downwards if $p > q$, and vice versa if $q < p$. The technical term for this behaviour is *drift*. A sample path for the case $p > q$ might look like:



This is 'unfair', because of the drift upwards, so we should suspect that the asymmetric random walk is not a martingale. In fact,

$$\mathbb{E}[S_n]=\sum_{i=1}^n\mathbb{E}[X_i]=\sum_{i=1}^n(p-q)=n(p-q), \tag{4.2}$$

whereas $\mathbb{E}[S_0]=0$. Thus, Lemma 3.3.6 confirms that S_n is not a martingale. However, the process

$$M_n=S_n-n(p-q) \tag{4.3}$$

is a martingale. The key is that the term $n(p-q)$ *compensates* for the drift and 'restores fairness'.

We'll now prove that (M_n) is a martingale. Since $X_i\in m\mathcal{F}_n$ for all $i\le n$, by Proposition 2.2.4 we have $S_n-n(p-q)\in m\mathcal{F}_n$. Since $|X_i|\le 1$ we have

$$|S_n-n(p-q)|\le|S_n|+n|p-q|\le n+n|p-q|$$

and hence M_n is bounded, so $M_n\in L^1$. Lastly,

$$\begin{aligned}\mathbb{E}[S_{n+1}-(n+1)(p-q)\,|\,\mathcal{F}_n]&=\mathbb{E}[S_{n+1}\,|\,\mathcal{F}_n]-(n+1)(p-q)\\&=\mathbb{E}[X_{n+1}\,|\,\mathcal{F}_n]+\mathbb{E}[S_n\,|\,\mathcal{F}_n]-(n+1)(p-q)\\&=\mathbb{E}[X_{n+1}]+S_n-(n+1)(p-q)\\&=(p-q)+S_n-(n+1)(p-q)\\&=S_n-n(p-q).\end{aligned}$$

Therefore $\mathbb{E}[M_{n+1}\,|\,\mathcal{F}_n]=M_n$, and (M_n) is a martingale.

4.2 Urn processes

Urn processes are 'balls in bags' processes. In the simplest kind of urn process, which we look at in this section, we have just a single urn that contains balls of two different colours, known as a *Pólya* urn.

At time 0, an urn contains 1 black ball and 1 red ball. Then, for each $n = 0, 1, 2, \dots$, we generate the state of the urn at time $n+1$ by doing the following:

1. Draw a ball from the urn, look at its colour, and return this ball to the urn.
2. Add a new ball of the same colour as the drawn ball.

So, at time n, there are $n+2$ balls in the urn, of which B_n+1 are black, where B_n is the number of black balls added into the urn before time n.

Let

$$M_n = \frac{B_n+1}{n+2}$$

be the fraction of balls in the urn that are black, at time n. Note that $M_n \in [0,1]$. Since (at least, on average) we are equally likely to add red balls as black balls, we might hope that M_n is 'fair' and that it is a martingale; which it is.

Set $(\mathcal{F}_n)$ to be the filtration generated by (B_n). Then we can calculate that

$$\begin{aligned}
\mathbb{E}[M_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}\left[M_{n+1}\mathbb{1}\{(n+1)^{th}\text{ draw is black}\}\,\middle|\,\mathcal{F}_n\right] + \mathbb{E}\left[M_{n+1}\mathbb{1}\{(n+1)^{th}\text{ draw is red}\}\,\middle|\,\mathcal{F}_n\right] \\
&= \mathbb{E}\left[\frac{B_n+2}{n+3}\mathbb{1}\{(n+1)^{th}\text{ draw is black}\}\,\middle|\,\mathcal{F}_n\right] + \mathbb{E}\left[\frac{B_n+1}{n+3}\mathbb{1}\{(n+1)^{th}\text{ draw is red}\}\,\middle|\,\mathcal{F}_n\right] \\
&= \frac{B_n+2}{n+3}\mathbb{E}\left[\mathbb{1}\{(n+1)^{th}\text{ draw is black}\}\,\middle|\,\mathcal{F}_n\right] + \frac{B_n+1}{n+3}\mathbb{E}\left[\mathbb{1}\{(n+1)^{th}\text{ draw is red}\}\,\middle|\,\mathcal{F}_n\right] \\
&= \frac{B_n+2}{n+3}\frac{B_n+1}{n+2} + \frac{B_n+1}{n+3}\left(1-\frac{B_n+1}{n+2}\right) \\
&= \frac{B_n^2+3B_n+2}{(n+2)(n+3)} + \frac{(n+2)B_n+(n+2)-(B_n^2+2B_n+1)}{(n+2)(n+3)} \\
&= \frac{(n+3)B_n+(n+3)}{(n+2)(n+3)} \\
&= \frac{B_n+1}{n+2} \\
&= M_n.
\end{aligned}$$

We have $M_n \in m\mathcal{F}_n$ and since $M_n \in [0,1]$ we have that $M_n \in L^1$. Hence (M_n) is a martingale.

4.2.1 On fairness

It is clear that the symmetric random walk is fair; at all times it is equally likely to move up as down. The asymmetric random walk is not fair, due to its drift (4.2), but once we compensate for drift in (4.3) we do still obtain a martingale.

Then urn process requires more careful thought. At first glance, we might wonder:

> *Suppose that the first draw is black. Then, at time $n=1$ we have two black balls and one red ball. So, the chance of drawing a black ball is now $\frac{2}{3}$. How is this fair?!*

To answer this question, let us make a number of points. Firstly, the quantity that is a martingale is M_n, the *fraction* of black balls in the urn (and not the absolute number of black balls!).

Secondly, suppose that the first draw is indeed black. So, at $n = 1$ we have 2 black and 1 red, giving a fraction $\frac{2}{3}$ of black. From here on, the expected fraction of black balls after the next draw is

$$\frac{2}{3} \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{2}{4} = \frac{6+2}{12} = \frac{2}{3}$$

which is of course equal to the fraction of black balls we had at $n = 1$. In this sense, the game is fair.

Lastly, note that it is equally likely that, on the first go, you'd pick out a red. So, starting from $n = 0$ and looking forwards, both colors have equally good chances of increasing their own numbers (in fact, by symmetry, the roles of red and black are interchangeable).

To sum up: in life there are different ways to think of 'fairness' – and what we need to do here is get a sense for precisely what kind of fairness martingales characterize. For example, the fact that M_n is a martingale does not mean that the proportions of red and black balls both remain $= \frac{1}{2}$ for all time (although you might think of $= \frac{1}{2}$ as a very strict type of 'fairness'). It just means that, when viewed in terms of M_n, there is no bias towards red of black inherent in the rules of the game.

4.3 A branching process

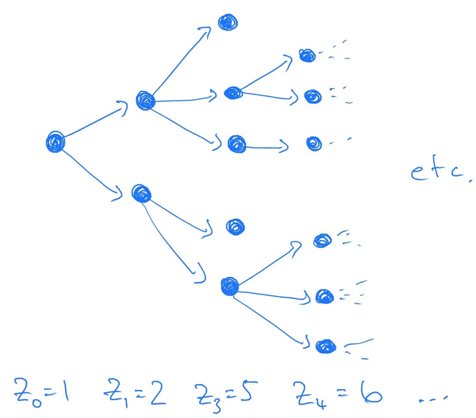
Branching processes are stochastic processes that model objects which divide up into a random number of copies of themselves. They are particularly important in mathematical biology (think of cell division, the tree of life, etc). We won't study any mathematical biology in this course, but we will look at one example of a branching process: the Galton-Waton process.

Let X_i^n, where $n, i \geq 1$, be i.i.d. nonnegative integer-valued random variables with common distribution G. Define a sequence (Z_n) by $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} X_1^{n+1} + \ldots + X_{Z_n}^{n+1}, & \text{if } Z_n > 0 \\ 0, & \text{if } Z_n = 0 \end{cases} \tag{4.4}$$

Then Z is the Galton-Watson process. Typically, we think of Z_n as representing the number of individuals in the n^{th} generation of some population, each of whom becomes the parent of an i.i.d. number of children in the $(n+1)^{th}$ generation. The (common) distribution G of the number of children X_i^n is known as the *offspring distribution.*

The name 'branching process' is best understood by drawing the process (Z_n) as a graph, in which arrows are drawn from parents to their children.



Note that if $Z_n = 0$ for some n, then for all $m > n$ we also have $Z_m = 0$.

Remark 4.3.1 ($\star$) *The Galton-Watson process takes its name from Francis Galton and Henry Watson, who in 1874 were concerned that Victorian aristocratic surnames were becoming extinct. They tried to model how many children people had, which is also how many times a surname was passed on, per family. This allowed them to use the process Z_n to predict whether a surname would die out (i.e. if $Z_n = 0$ for some n) or become widespread (i.e. $Z_n \to \infty$).*

(Since then, the Galton-Watson process has found more important uses.)

Let $\mu = \mathbb{E}[G]$, and let $\mathcal{F}_n = \sigma(X_{m,i}\,;\, i \in \mathbb{N}, m \leq n)$. In general, Z_n is not a martingale because

$$\begin{aligned}
\mathbb{E}[Z_{n+1}] &= \mathbb{E}\left[X_1^{n+1} + \ldots + X_{Z_n}^{n+1}\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right) \mathbb{1}\{Z_n = k\}\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right)\right] \mathbb{E}\left[\mathbb{1}\{Z_n = k\}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=1}^{\infty} \left(\mathbb{E}\left[X_1^{n+1}\right] + \ldots + \mathbb{E}\left[X_k^{n+1}\right]\right) \mathbb{P}\left[Z_n = k\right] \\
&= \sum_{k=1}^{\infty} k\mu \mathbb{P}[Z_n = k] \\
&= \mu \sum_{k=1}^{\infty} k \mathbb{P}[Z_n = k] \\
&= \mu\, \mathbb{E}[Z_n].
\end{aligned} \tag{4.5}
$$

Lemma 3.3.6 tells us that if (M_n) is a martingale that $\mathbb{E}[M_n] = \mathbb{E}[M_{n+1}]$. But, if $\mu < 1$ we see that $\mathbb{E}[Z_{n+1}] < \mathbb{E}[Z_n]$ (downwards drift) and if $\mu > 1$ then $\mathbb{E}[Z_{n+1}] > \mathbb{E}[Z_n]$ (upwards drift).

However, much like with the asymmetric random walk, we can compensate for the drift and obtain a martingale. More precisely, we will show that

$$
M_n = \frac{Z_n}{\mu^n}
$$

is a martingale.

We have $M_0 = 1 \in m\mathcal{F}_0$, and if $M_n \in \mathcal{F}_n$ then from (4.4) we have that $M_{n+1} \in m\mathcal{F}_{n+1}$. Hence, by induction $M_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. From (4.5), we have $\mathbb{E}[Z_{n+1}] = \mu\mathbb{E}[Z_n]$ so as $\mathbb{E}[Z_n] = \mu^n$ for all n. Hence $\mathbb{E}[M_n] = 1$ and $M_n \in L^1$. Lastly,

$$
\begin{aligned}
\mathbb{E}[Z_{n+1} \,|\, \mathcal{F}_n] &= \sum_{k=1}^{\infty} \mathbb{E}[Z_{n+1} \mathbb{1}\{Z_n = k\} \,|\, \mathcal{F}_n] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\left(X_1^{n+1} + \ldots + X_k^{n+1}\right) \mathbb{1}\{Z_n = k\} \,|\, \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} \mathbb{1}\{Z_n = k\} \mathbb{E}\left[X_1^{n+1} + \ldots + X_k^{n+1} \,|\, \mathcal{F}_n\right] \\
&= \sum_{k=1}^{\infty} k\mu \mathbb{1}\{Z_n = k\} \\
&= \mu Z_n.
\end{aligned}
$$

Here we use that Z_n is $\mathcal{F}_n$ measurable to take out what is known, and then use that X_i^{n+1} is independent of $\mathcal{F}_n$. Hence, $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$, as required.

4.4 Other stochastic processes

The world of stochastic processes, like the physical world that they try to model, is many and varied. We can make more general kinds of random walk (and urn/branching processes) by allowing more complex rules for what should happen on each new time step. Those of you who have taken MAS275 will have seen renewal processes and Markov chains, which are two more important types of stochastic process. There are stochastic processes to model objects that coalesce together, objects that move around in space, objects that avoid one another, etc, etc.

Most (but not all) types of stochastic process have connections to martingales. The reason for making these connections is that by using martingales it is possible to extract information about the behaviour of stochastic a process – we will see some examples of how this can be done in Chapters 7 and 8.

4.5 Exercises

On stochastic processes

4.1 Let $S_n = \sum_{i=1}^n X_i$ be the symmetric random walk from Section 4.1.1 and define $Z_n = e^{S_n}$. Show that Z_n is a submartingale and that

$$M_n = \left(\frac{2}{e + \frac{1}{e}}\right)^n Z_n$$

is a martingale.

4.2 Let (X_i) be a sequence of identically distributed random variables with common distribution

$$X_i = \begin{cases} a & \text{with probability } p_a \\ -b & \text{with probability } p_b = 1 - p_a. \end{cases}$$

where $0 \leq a, b$. Let $S_n = \sum_{i=1}^n X_i$. Under what conditions on a, b, p_a, p_b is (S_n) a martingale?

4.3 Let (X_i) be an i.i.d. sequence of random variables such that $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = 1$. Define a stochastic process S_n by setting $S_0 = 1$ and

$$S_{n+1} = \begin{cases} S_n + X_{n+1} & \text{if } S_n > 0, \\ 1 & \text{if } S_n = 0. \end{cases}$$

That is, S_n behaves like a symmetric random walk but, whenever it hits zero, on the next time step it is 'reflected' back to 1. Let $L_n = \sum_{i=0}^{n-1} \mathbb{1}\{S_i = 0\}$ be the number of time steps, before time n, at which S_n is zero. Show that

$$\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] = S_n + \mathbb{1}\{S_n = 0\}$$

and hence show that $S_n - L_n$ is a martingale.

4.4 Consider an urn that may contain balls of three colours: red, blue and green. Initially the urn contains one ball of each colour. Then, at each step of time $n = 1, 2, \ldots$ we draw a ball from the urn. We place the drawn ball back into the urn and add an additional ball of the same colour.

Let (M_n) be the proportion of balls that are red. Show that (M_n) is a martingale.

4.5 Let $S_n = \sum_{i=1}^n X_i$ be the symmetric random walk from Section 4.1.1. State, with proof, which of the following processes are martingales:

$$\text{(i) } S_n^2 + n \qquad \text{(ii) } S_n^2 - n \qquad \text{(iii) } \frac{S_n}{n}$$

Which of the above are submartingales?

Challenge questions

4.6 Let (S_n) be the symmetric random walk from Section 4.1.1. Prove that there is no deterministic function $f : \mathbb{N} \to \mathbb{R}$ such that $S_n^3 - f(n)$ is a martingale.

Chapter 5

The binomial model

We now return to financial mathematics. We will extend the one-period model from Chapter 1 and discover a surprising connection between arbitrage and martingales.

5.1 Arbitrage in the one-period model

Let us recall the one-period market from Section 1.2. We have two commodities, cash and stock. Cash earns interest at rate r, so:

- If we hold x units of cash at time 0, they become worth $x(1+r)$ at time 1.

At time $t=0$, a single unit of stock is worth s units of cash. At time 1, the value of a unit of stock changes to

$$S_1 = \begin{cases} sd & \text{with probability } p_d, \\ su & \text{with probability } p_u, \end{cases}$$

where $p_u + p_d = 1$.

Note that roles of u and d are interchangeable – we would get the same model if we swapped the values of u and d (and p_u and p_d to match). So, we lose nothing by assuming that $d < u$.

The price of our stock changes as follows:

- If we hold y units of stock, worth ys, at time 0, they become worth yS_1 at time 1.

Recall that we can borrow cash from the bank (provided we pay it back with interest at rate r, at some later time) and that we can borrow stock from the stockbroker (provided we give the same number of units of stock back, at some later time). Thus, x and y are allowed to be negative, with the meaning that we have borrowed.

Recall also that we use the term portfolio for the amount of cash/stock that we hold at some time. We can formalize this: A **portfolio** is a pair $h = (x, y) \in \mathbb{R}^2$, where x is the amount of cash and y is the number of (units of) of stock.

Definition 5.1.1 *The* ***value process*** *or* ***price process*** *of the portfolio $h = (x, y)$ is the process V^h given by*

$$\begin{aligned} V_0^h &= x + ys \\ V_1^h &= x(1+r) + yS_1. \end{aligned}$$

We can also formalize the idea of arbitrage. A portfolio is an arbitrage if it makes money for free:

Definition 5.1.2 *A portfolio $h = (x, y)$ is said to be an* ***arbitrage possibility*** *if:*

$$
\begin{aligned}
V_0^h &= 0 \\
\mathbb{P}[V_1^h \geq 0] &= 1. \\
\mathbb{P}[V_1^h > 0] &> 0.
\end{aligned}
$$

We say that a market is ***arbitrage free*** *if there do not exist any arbitrage possibilities.*

It is possible to characterize exactly when the one-period market is arbitrage free. In fact, we have already done most of the work in **1.3**.

Proposition 5.1.3 *The one-period market is arbitrage free if and only if $d < 1 + r < u$.*

PROOF: $(\Rightarrow)$: Recall that we assume $d < u$. Hence, if $d < 1 + r < u$ fails then either $1 + r \leq d < u$ or $d < u \leq 1 + r$. In both cases, we will construct an arbitrage possibility.

In the case $1 + r \leq d < u$ we use the portfolio $h = (-s, 1)$ which has $V_0^h = 0$ and

$$V_1^h = -s(1 + r) + S_1 \geq s(-(1 + r) + d) \geq 0,$$

hence $\mathbb{P}[V_1^h \geq 0] = 1$. Further, with probability $p_u > 0$ we have $S_1 = su$, which means $V_1^h > s(-(1 + r) + d) \geq 0$. Hence $\mathbb{P}[V_1^h > 0] > 0$. Thus, h is an arbitrage possibility.

If $0 < d < u \leq 1 + r$ then we use the portfolio $h' = (s, -1)$, which has $V_0^{h'} = 0$ and

$$V_1^{h'} = s(1 + r) - S_1 \geq s(1 + r - u) \geq 0,$$

hence $\mathbb{P}[V_1^{h'} \geq 0] = 1$. Further, with probability $p_d > 0$ we have $S_1 = sd$, which means $V_1^{h'} > s(-(1 + r) + u) \geq 0$. Hence $\mathbb{P}[V_1^{h'} > 0] > 0$. Thus, h' is also an arbitrage possibility.

Remark 5.1.4 *In both cases, at time 0 we borrow whichever commodity (cash or stock) will grow slowest in value, immediately sell it and use the proceeds to buy the other, which we know will grow faster in value. Then we wait; at time 1 we own the commodity has grown fastest in value, so we sell it, repay our debt and have some profit left over.*

$(\Leftarrow)$: Now, assume that $d < 1 + r < u$. We need to show that no arbitrage is possible. To do so, we will show that if a portfolio has $V_0^h = 0$ and $V_1^h \geq 0$ then it also has $V_1^h = 0$.

So, let $h = (x, y)$ be a portfolio such that $V_0^h = 0$ and $V_1^h \geq 0$. We have

$$V_0^h = x + ys = 0.$$

The value of h at time 1 is

$$V_1^h = x(1 + r) + ysZ.$$

Using that $x = -ys$, we have

$$V_1^h = \begin{cases} ys(u - (1 + r)) & \text{if } Z = u, \\ ys(d - (1 + r)) & \text{if } Z = d. \end{cases} \tag{5.1}$$

Since $\mathbb{P}[V_1^h \geq 0] = 1$ this means that both (a) $ys(u - (1 + r)) \geq 0$ and (b) $ys(d - (1 + r)) \geq 0$. If $y < 0$ then we contradict (a) because $1 + r < u$. If $y > 0$ then we contradict (b) because $d < 1 + r$. So the only option left is that $y = 0$, in which case $V_0^h = V_1^h = 0$. ■

5.1.1 Expectation regained

In Proposition 5.1.3 we showed that our one period model was free of arbitrage if and only if

$$d < 1 + r < u.$$

This condition is very natural: it means that sometimes the stock will outperform cash and sometimes cash will outperform the stock. Without this condition it is intuitively clear that our market would be a bad model. From that point of view, Proposition 5.1.3 is encouraging since it confirms the importance of (no) arbitrage.

However, it turns out that there is more to the condition $d < 1 + r < u$, which we now explore. It is equivalent to asking that there exists $q_u, q_d \in (0, 1)$ such that both

$$q_u + q_d = 1 \quad \text{and} \quad 1 + r = uq_u + dq_d. \tag{5.2}$$

In words, (5.2) says that $1 + r$ is a weighted average of d and u. We could solve these two equations to see that

$$q_u = \frac{(1 + r) - d}{u - d}, \quad q_d = \frac{u - (1 + r)}{u - d}. \tag{5.3}$$

Now, here is the key: we can think of the weights q_u and q_d as probabilities. Let's pretend that we live in a different world, where a single unit of stock, worth $S_0 = s$ at time 0, changes value to become worth

$$S_1 = \begin{cases} sd & \text{with probability } q_d, \\ su & \text{with probability } q_u. \end{cases}$$

We have altered – the technical term is **tilted** – the probabilities from their old values p_d, p_u to new values q_d, q_u. Let's call this new world $\mathbb{Q}$, by which we mean that $\mathbb{Q}$ is our new probability measure: $\mathbb{Q}[S_1 = sd] = q_d$ and $\mathbb{Q}[S_1 = su] = q_u$. This is often called the **risk-neutral world**, and q_u, q_d are known as the **risk-neutral probabilities**[1].

Since $\mathbb{Q}$ is a probability measure, we can use it to take expectations. We use $\mathbb{E}^{\mathbb{P}}$ and $\mathbb{E}^{\mathbb{Q}}$ to make it clear if we are taking expectations using $\mathbb{P}$ or $\mathbb{Q}$.

We have

$$\begin{aligned} \frac{1}{1 + r}\mathbb{E}^{\mathbb{Q}}[S_1] &= \frac{1}{1 + r}\Big(su\mathbb{Q}[S_1 = su] + sd\mathbb{Q}[S_1 = sd]\Big) \\ &= \frac{1}{1 + r}(s)(uq_u + dq_d) \\ &= s. \end{aligned}$$

The price of the stock at time 0 is $S_0 = s$. To sum up, we have shown that the price S_t of a unit of stock at time t satisfies

$$S_0 = \frac{1}{1 + r}\mathbb{E}^{\mathbb{Q}}[S_1]. \tag{5.4}$$

This is a formula that is very well known to economists. It gives the stock price today ($t = 0$) as the expectation under $\mathbb{Q}$ of the stock price tomorrow ($t = 1$), **discounted** by the rate $1 + r$ at which it would earn interest.

Equation (5.4) is our first example of a 'risk-neutral valuation' formula. Recall that we pointed out in Chapter 1 that we should not use $\mathbb{E}^{\mathbb{P}}$ and 'expected value' prices. A possible

[1] We will discuss the reason for the name 'risk-neutral' later. It is standard terminology in the world of stocks and shares.

cause of confusion is that (5.4) *does* correctly calculate the value (i.e. price) of a single unit of stock by taking an expectation. The point is that we (1) use $\mathbb{E}^{\mathbb{Q}}$ rather than $\mathbb{E}^{\mathbb{P}}$ and (2) then discount according to the interest rate. We will see, in the next section, that these two steps are the correct way to go about arbitrage free pricing in general.

Moreover, in Section 5.4 we will extend our model to have multiple time steps. Then the expectation in (5.4) will lead us to martingales.

5.2 Hedging in the one-period model

We saw in Section 1.3 that the 'no arbitrage' assumption could force *some* prices to take particular values. It is not immediately obvious if the absence of arbitrage forces a unique value for *every* price; we will show in this section that it does.

First, let us write down exactly what it is that we need to price.

Definition 5.2.1 *A* ***contingent claim*** *is any random variable of the form $X = \Phi(S_1)$, where Φ is a deterministic function.*

The function Φ is sometimes known as the contract function. One example of a contingent claim is a **forward contract**, in which the holder promises to buy a unit of stock at time 1 for a fixed price K. In this case the contingent claim would be

$$\Phi(S_1) = S_1 - K,$$

the value of a unit of stock at time 1 minus the price paid for it. We will see many other examples in the course. Here is another.

Example 5.2.2 *A* ***European call option*** *gives its holder the right (but not the obligation) to buy, at time 1, a single unit of stock for a fixed price K that is agreed at time 0. As for futures, K is known as the strike price.*

Suppose we hold a European call option at time 1. Then, if $S_1 > K$, we could exercise our right to buy a unit of stock at price K, immediately sell the stock for S_1 and consequently earn $S_1 - K > 0$ in cash. Alternatively if $S_1 \leq K$ then our option is worthless.

Since S_1 is equal to either to either su or sd, the only interesting case is when $sd < K < su$. In this case, the contingent claim for our European call option is

$$\Phi(S_1) = \begin{cases} su - K & \text{if } S_1 = su \\ 0 & \text{if } S_1 = sd. \end{cases} \tag{5.5}$$

In the first case our right to buy is worth exercising; in the second case it is not. A simpler way to write this contingent claim is

$$\Phi(S_1) = \max(S_1 - K, 0). \tag{5.6}$$

In general, given any contract, we can work out its contingent claim. We therefore plan to find a general way of pricing contingent claims. In Section 1.3 we relied on finding specific trading strategies to determine prices (one, from the point of view of the buyer, that gave an upper bound and one, from the point of view of the seller, to give a lower bound). Our first step in this section is to find a general way of constructing trading strategies.

Definition 5.2.3 *We say that a portfolio h is a* ***replicating portfolio*** *or* ***hedging portfolio*** *for the contingent claim $\Phi(S_1)$ if $V_1^h = \Phi(S_1)$.*

The process of finding a replicating portfolio is known simply as replicating or hedging. The above definition means that, if we hold the portfolio h at time 0, then at time 1 it will have precisely the same value as the contingent claim $\Phi(S_1)$. Therefore, since we assume our model is free of arbitrage:

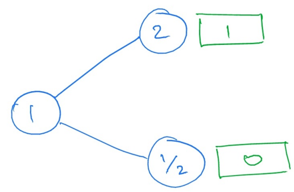
> If a contingent claim $\Phi(S_1)$ has a replicating portfolio h, then the price of the $\Phi(S_1)$ at time 0 must be equal to the value of h at time 0.

We say that a market is **complete** if every contingent claim can be replicated. Therefore, if the market is complete, we can price any contingent claim.

Example 5.2.4 *Suppose that $s = 1, d = \frac{1}{2}, u = 2$ and $r = \frac{1}{4}$, and that we are looking at the contingent claim*

$$\Phi(S_1) = \begin{cases} 1 & \text{if } S_1 = su, \\ 0 & \text{if } S_1 = sd. \end{cases}$$

We can represent this situation as a tree, with a branch for each possible movement of the stock, and the resulting value of our contingent claim written in a square box.



Suppose that we wish to replicate $\Phi(S_1)$. That is, we need a portfolio $h = (x, y)$ such that $V_1^h = \Phi(S_1)$:

$$(1 + \tfrac{1}{4})x + 2y = 1$$
$$(1 + \tfrac{1}{4})x + \tfrac{1}{2}y = 0.$$

This is a pair of linear equations that we can solve. The solution (which is left for you to check) is $x = \frac{-4}{15}$, $y = \frac{2}{3}$. Hence the price of our contingent claim $\Phi(S_1)$ at time 0 is $V_0^h = \frac{-4}{15} + 1 \cdot \frac{2}{3} = \frac{2}{5}$.

Let us now take an arbitrary contingent claim $\Phi(S_1)$ and see if we can replicate it. This would mean finding a portfolio h such that the value V_1^h of the portfolio at time 1 is $\Phi(S_1)$:

$$V_1^h = \begin{cases} \Phi(su) & \text{if } S_1 = su, \\ \Phi(sd) & \text{if } S_1 = sd. \end{cases}$$

By (5.1), if we write $h = (x, y)$ then we need

$$(1 + r)x + suy = \Phi(su)$$
$$(1 + r)x + sdy = \Phi(sd),$$

which is just a pair of linear equations to solve for (x, y). In matrix form,

$$\begin{pmatrix} 1+r & su \\ 1+r & sd \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \Phi(su) \\ \Phi(sd) \end{pmatrix}. \tag{5.7}$$

A unique solution exists when the determinant is non-zero, that is when $(1+r)u-(1+r)d \neq 0$, or equivalently when $u \neq d$. So, in this case, we can find a replicating portfolio for any contingent claim.

It is an assumption of the model that $d \leq u$, so we have that our one-period model is complete if $d < u$. Therefore:

Proposition 5.2.5 *If the one-period model is arbitrage free then it is complete.*

And, in this case, we can solve (5.7) to get

$$\begin{aligned} x &= \frac{1}{1+r}\frac{u\Phi(sd)-d\Phi(su)}{u-d}, \\ y &= \frac{1}{s}\frac{\Phi(su)-\Phi(sd)}{u-d}. \end{aligned} \tag{5.8}$$

which tells us that the price of $\Phi(S_1)$ at time 0 should be

$$\begin{aligned} V_0^h &= x + sy \\ &= \frac{1}{1+r}\left(\frac{(1+r)-d}{u-d}\Phi(su) + \frac{u-(1+r)}{u-d}\Phi(sd)\right) \\ &= \frac{1}{1+r}\left(q_u\Phi(su) + q_d\Phi(sd)\right) \\ &= \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[\Phi(S_1)]. \end{aligned}$$

Hence, the value (and therefore, the price) of $\Phi(S_1)$ at time 0, is given by

$$V_0^h = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[\Phi(S_1)]. \tag{5.9}$$

The formula (5.9) is known as the **risk-neutral valuation formula**. It says that to find the price of $\Phi(S_1)$ at time 0 we should take its expectation according to $\mathbb{Q}$, and then discount one time step worth of interest i.e. divide by $1+r$. It is a very powerful tool, since it allows us to price any contingent claim.

Note the similarity of (5.9) to (5.4). In fact, (5.4) is a special case of (5.9), namely the case where $\Phi(S_1) = S_1$ i.e. pricing the contingent claim corresponding to being given a single unit of stock.

To sum up:

Proposition 5.2.6 *Let $\Phi(S_1)$ be a contingent claim. Then the (unique) replicating portfolio $h = (x, y)$ for $\Phi(S_1)$ can be found by solving $V_1^h = \Phi(S_1)$, which can be written as a pair of linear equations:*

$$\begin{aligned} (1+r)x + suy &= \Phi(su) \\ (1+r)x + sdy &= \Phi(sd). \end{aligned}$$

The general solution is (5.8)*. The value (and hence, the price) of $\Phi(S_1)$ at time* 0 *is*

$$V_0^h = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\Phi(S_1)\right].$$

For example, we can now both price and hedge the European call option.

Example 5.2.7 *We found the contingent claim of a European call option with strike price $K \in (sd, su)$ in* (5.5)*. By the first part of Proposition 5.2.6, to find a replicating portfolio $h = (x, y)$ we must solve $V_1^h = \Phi(S_1)$, which is*

$$\begin{aligned} (1+r)x + suy &= su - K \\ (1+r)x + sdy &= 0. \end{aligned}$$

This has the solution (again, left for you to check) $x = \frac{sd(K-su)}{(1+r)(su-sd)}, y = \frac{su-K}{su-sd}$. By the second *part of Proposition 5.2.6 the value of the European call option at time* 0 *is*

$$\begin{aligned}\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\Phi(S_1)\right] &= \frac{1}{1+r}\left(q_u(su-K)+q_d(0)\right)\\ &= \frac{1}{1+r}\frac{(1+r)-d}{u-d}(su-K).\end{aligned}$$

5.3 Types of financial derivative

A contract that specifies that buying/selling will occur, now or in the future, is known as a **financial derivative**, or simply derivative. Financial derivatives that give a choice between two options are often known simply as **options**.

Here we collect together the various types of financial derivatives that we have mentioned in previous sections. This necessitates some repetition. We also include some new types of option that have not been mentioned before, but will appear in future chapters. As before, we use the term **strike price** to refer to a fixed price K that is agreed at time 0 (and paid at time 1).

- A **future** or **future contract** is the obligation to be given a single unit of stock at time 1.
- A **forward** or **forward contract** is the obligation to buy a single unit of stock at time 1 for a strike price K.
- A **European call option** is the right, but not the obligation, to buy a single unit of stock at time 1 for a strike price K.
- A **European put option** is the right, but not the obligation, to sell a single unit of stock at time 1 for a strike price K.

You are expected to remember these definitions!

There are many other types of financial derivative; we'll look at more examples later in the course.

5.4 The binomial model

Let us step back and examine our progress, for a moment. We now know about as much about one-period model as there is to know. It is time to move onto to a more complicated (and more realistic) model. The one-period model is unsatisfactory in two main respects:

1. The one-period model has only a single step of time.

2. The stock price process (S_t) is too simplistic.

We'll start to address the first of these points now. The second point waits until the second semester of the course.

Adding multiple time steps to our model will make use of the theory we developed in Chapters 2 and 3. It will also reveal a surprising connection between arbitrage and martingales.

The **binomial model** has time points $t = 0, 1, 2, \ldots, T$. Inside each time step, we have a single step of the one-period model. This means that cash earns interest at rate r:

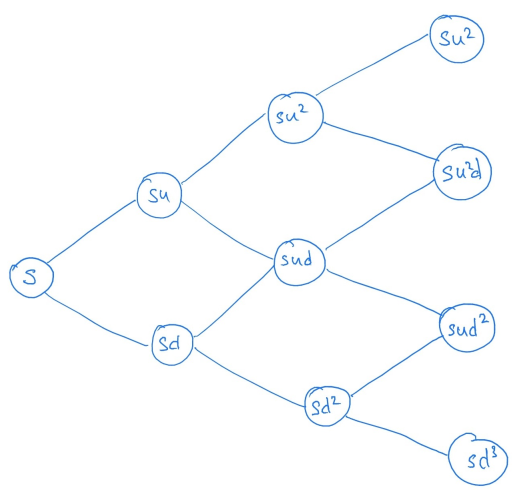
- If we hold x units of cash at time t, it will become worth $x(1 + r)$ (in cash) at time $t + 1$.

For our stock, we'll have to think a little harder. In a single time step, the value of our stock is multiplied by a random variable Z with distribution (**??**). We now have several time steps. For each time step we'll use a new *independent* Z. So, let $(Z_t)_{t=1}^{T}$ be a sequence of i.i.d. random variables each with the distribution of Z.

- The value of a single unit of stock at time t is given by

$$
\begin{aligned}
S_0 &= s, \\
S_t &= Z_t S_{t-1}.
\end{aligned}
$$

We can illustrate the process (S_t) using a tree-like diagram:



Note that the tree is recombining, in the sense that a move up (by u) followed by a move down (by d) has the same outcome as a move down followed by a move up. It's like a random walk, except we multiply instead of add (recall exercise **4.1**).

Remark 5.4.1 *The one-period model is simply the $T = 1$ case of the binomial model.*

5.5 Portfolios, arbitrage and martingales

Since we now have multiple time steps, we can exchange cash for stock (and vice versa) at all times $t = 0, 1, \dots, T-1$. We need to expand our idea of a portfolio to allow for this.

The filtration corresponding to the information available to a buyer/seller in the binomial model is

$$\mathcal{F}_t = \sigma(Z_1, Z_2, \dots, Z_t).$$

In words, the information in $\mathcal{F}_t$ contains changes in the stock price up to and including at time t. This means that, $S_0, S_1, \dots, S_t$ are all $\mathcal{F}_t$ measurable, but S_{t+1} is not $\mathcal{F}_t$ measurable.

When we choose how much stock/cash to buy/sell at time $t-1$, we do so without knowing how the stock price will change during $t-1 \mapsto t$. So we must do so using information only from $\mathcal{F}_{t-1}$. This makes the following general definition useful.

Definition 5.5.1 *A stochastic process (S_t) is said to be previsible if $S_t \in \mathcal{F}_{t-1}$ for all $t \in \mathbb{N}$.*

We now have enough terminology to define the strategies that are available to participants in the binomial market.

Definition 5.5.2 *A **portfolio strategy** is a stochastic process*

$$h_t = (x_t, y_t)$$

for $t = 0, 1, 2, \dots, T$, such that (h_t) is previsible.

The interpretation is that x_t is the amount of cash, and y_t the amount of stock, that we hold during the time step $t-1 \mapsto t$. Requiring that (h_t) is previsible means that h_t is $\mathcal{F}_{t-1}$ measurable. That is, we make our choice of how much cash and stock to hold during $t-1 \mapsto t$ based on knowing the value of $S_0, S_1, \dots, S_{t-1}$, but *without* knowing S_t. This is realistic.

Definition 5.5.3 *The **value process** of the portfolio strategy $h = (h_t)_{t=1}^T$ is the stochastic process (V_t) given by*

$$\begin{aligned} V_0^h &= x_0 + y_0 S_0, \\ V_t^h &= x_t(1+r) + y_t S_t, \end{aligned}$$

for $t = 1, 2, \dots, T$.

At $t = 0$, V_0^h is the value of the portfolio h_0. For $t \geq 1$, V_t^h is the value of the portfolio (x_t, y_t) at time t, after the change in value of cash/stock that occurs during $t-1 \mapsto t$. The value process is measurable but it is not previsible.

We will be especially interested in portfolio strategies that require an initial investment at time 0 but, at later times $t \geq 1, 2, \dots, T-1$, any changes in the amount of stock/cash held will pay for itself. We capture such portfolio strategies in the following definition.

Definition 5.5.4 *A portfolio strategy $h_t = (x_t, y_t)$ is said to be **self-financing** if*

$$V_{t-1}^h = x_t + y_t S_{t-1}.$$

for $t = 1, 2, \dots, T$.

This means that the value of the portfolio at time $t-1$ is equal to the value (at time $t-1$) of the stock/cash that is held in between times $t-1 \mapsto t$. In other words, in a self-financing portfolio at the times $t = 1, 2, \ldots$ we can swap our stocks for shares (and vice versa) according to whatever the stock price turns out to be, but that is all we can do.

Lastly, our idea of arbitrage must also be upgraded to handle multiple time steps.

Definition 5.5.5 *We say that a portfolio strategy (h_t) is an **arbitrage possibility** if it is self-financing and satisfies*

$$\begin{aligned}V_0^h &= 0\\ \mathbb{P}[V_T^h \geq 0] &= 1.\\ \mathbb{P}[V_T^h > 0] &> 0.\end{aligned}$$

In words, an arbitrage possibility requires that we invest nothing at times $t = 0, 1, \ldots, T-1$, but which gives us a positive probability of earning something at time T, with no risk at all of actually losing money.

It's natural to ask when the binomial model is arbitrage free. Happily, the condition turns out to be the same as for the one-period model.

Proposition 5.5.6 *The binomial model is arbitrage free if and only if $d < 1 + r < u$.*

The proof is quite similar to the argument for the one-period model, but involves more technical calculations and (for this reason) we don't include it as part of the course.

Recall the risk-neutral probabilities from (5.3). In the one-period model, we use them to define the **risk-neutral** world $\mathbb{Q}$, in which on each time step the stock price moves up (by u) with probability q_u, or down (by d) with probability q_d. This provides a connection to martingales:

Proposition 5.5.7 *If $d < 1 + r < u$, then under the probability measure $\mathbb{Q}$, the process*

$$M_t = \frac{1}{(1+r)^t} S_t$$

is a martingale, with respect to the filtration $(\mathcal{F}_t)$.

Proof: We have commented above that $S_t \in m\mathcal{F}_t$, and we also have $d^t S_0 \leq S_t \leq u^t S_0$, so S_t is bounded and hence $S_t \in L^1$. Hence also $M_t \in m\mathcal{F}_t$ and $M_t \in L^1$. It remains to show that

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[M_{t+1} \,|\, \mathcal{F}_t] &= \mathbb{E}^{\mathbb{Q}}\left[M_{t+1}\mathbb{1}_{\{Z_{t+1}=u\}} + M_{t+1}\mathbb{1}_{\{Z_{t+1}=d\}} \,|\, \mathcal{F}_t\right]\\
&= \mathbb{E}^{\mathbb{Q}}\left[\frac{uS_t}{(1+r)^{t+1}}\mathbb{1}_{\{Z_{t+1}=u\}} + \frac{dS_t}{(1+r)^{t+1}}\mathbb{1}_{\{Z_{t+1}=d\}} \,|\, \mathcal{F}_t\right]\\
&= \frac{S_t}{(1+r)^{t+1}}\left(u\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=u\}} \,|\, \mathcal{F}_t\right] + d\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=d\}} \,|\, \mathcal{F}_t\right]\right)\\
&= \frac{S_t}{(1+r)^{t+1}}\left(u\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=u\}}\right] + d\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{\{Z_{t+1}=d\}}\right]\right)\\
&= \frac{S_t}{(1+r)^{t+1}}\left(u\mathbb{Q}[Z_{t+1}=u] + d\mathbb{Q}\left[Z_{t+1}=d\right]\right)\\
&= \frac{S_t}{(1+r)^{t+1}}(uq_u + dq_d)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{S_t}{(1+r)^{t+1}}(1+r) \\
&= M_t.
\end{aligned}
$$

Here, from the second to third line we take out what is known, using that $S_t \in m\mathcal{F}_t$. To deduce the third line we use linearity, and to deduce the fourth line we use that Z_{t+1} is independent of $\mathcal{F}_t$. Lastly, we recall from (5.2) that $uq_u + dq_d = 1 + r$. Hence, (M_t) is a martingale with respect to the filtration $\mathcal{F}_t$, in the risk-neutral world $\mathbb{Q}$. ■

Remark 5.5.8 *Using Lemma 3.3.6 we have* $\mathbb{E}^{\mathbb{Q}}[M_0] = \mathbb{E}^{\mathbb{Q}}[M_1]$*, which states that* $S_0 = \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}[S_1]$*. This is precisely* (5.4)

5.5.1 ($\star$) First Fundamental Theorem

In fact, the converse statement to Proposition 5.5.7 is also true; a probability measure $\mathbb{Q}$ under which M_t is a martingale exists *if and only if* $d < 1 + r < q$. Proof of this statement, which essentially involves working through many the arguments we've already seen in reverse, is beyond the scope of our course. Combined with Proposition 5.5.6, we obtain:

Theorem 5.5.9 (First Fundamental Theorem of Asset Pricing) *The following conditions are all equivalent:*

1. $d < 1 + r < u$*. (In words: sometimes cash out-performs stock, and sometimes it doesn't.)*
2. *No arbitrage possibilities exist.*
3. *There exists a probability measure* $\mathbb{Q}$ *such that* $M_t = \frac{S_t}{(1+r)^t}$ *is a* $\mathcal{F}_t$ *martingale under* $\mathbb{Q}$*.*

With this theorem in mind, the risk-neutral probability measure $\mathbb{Q}$ is sometimes known as the martingale measure, and q_d and q_d as the martingale probabilities.

Theorem 5.5.9 is important because it connects arbitrage to martingales in a way that, as it turns out, is true in very general models; we might include exchange rates between currencies, a more complex stock price process, transaction taxes, etc, and (equivalents of) the first fundamental theorem still turn out to hold. As a result, the first fundamental theorem provides an important piece of intuition for understanding financial markets.

5.6 Hedging

We can adapt the derivatives from Section 5.3 to the binomial model, by simply replacing time 1 with time T. For example, in the binomial model a forward contract is the obligation to buy a single unit of stock at time T for a strike price K that is agreed at time 0.

Definition 5.6.1 *A* ***contingent claim*** *is a random variable of the form* $\Phi(S_T)$*, where* $\Phi : \mathbb{R} \to \mathbb{R}$ *is a deterministic function.*

For a forward contract, the contingent claim would be $\Phi(S_T) = S_T - K$.

Definition 5.6.2 *We say that a portfolio strategy* $h = (h_t)_{t=1}^T$ *is a* ***replicating portfolio*** *or* ***hedging strategy*** *for the contingent claim* $\Phi(S_T)$ *if* $V_T^h = \Phi(S_T)$.

These match the definitions for the one-period model, except we now care about the value of the asset at time T (instead of time 1). We will shortly look at how to find replicating portfolios.

As in the one-period model, the binomial model is said to be **complete** if every contingent claim can be replicated. Further, as in the one-period model, the binomial model is complete if and only if it is free of arbitrage. With this in mind, for the rest of this section we assume that

$$d < 1 + r < u.$$

Lastly, as in the one-period model, our assumption that there is no arbitrage means that:

> If a contingent claim $\Phi(S_T)$ has a replicating strategy $h = (h_t)_{t=1}^T$, then the price of the $\Phi(S_T)$ at time 0 must be equal to the value of h_0.

Now, let us end this chapter by showing how to compute prices and replicating portfolios in the binomial model. We already know how to do this in the one-period model, see Example 5.2.4. We could do it in full generality (as we did in (5.7) for the one-period model) but this would involve lots of indices and look rather messy. Instead, we'll work through a practical example that makes the general strategy clear.

Let us take $T = 3$ and set

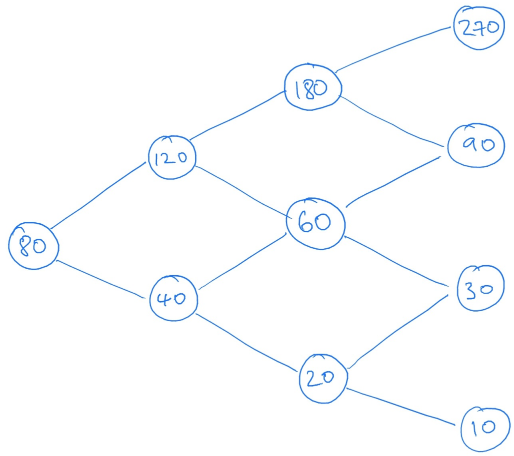
$$S_0 = 80, \quad u = 1.5, \quad d = 0.5, \quad p_u = 0.6, \quad p_d = 0.4.$$

To make the calculations easier, we'll also take our interest rate to be $r = 0$. We'll price a European call option with strike price $K = 80$. The contingent claim for this option, which is

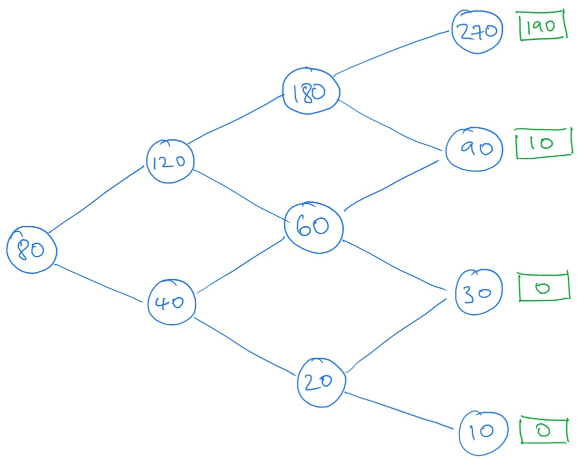
$$\Phi(S_T) = \max(S_T - K, 0). \tag{5.10}$$

STEP 1 is to work our the risk-neutral probabilities. From (5.3), these are $q_u = \frac{1+0-0.5}{1.5-0.5} = 0.5$ and $q_d = 1 - q_u = 0.5$.

STEP 2 is to write down the tree of possible values that the stock can take during time $t = 0, 1, 2, 3$. This looks like



We then work out, at each of the nodes corresponding to time $T = 3$, what the value of our contingent claim (5.10) would be if this node were reached. We write these values in square boxes:



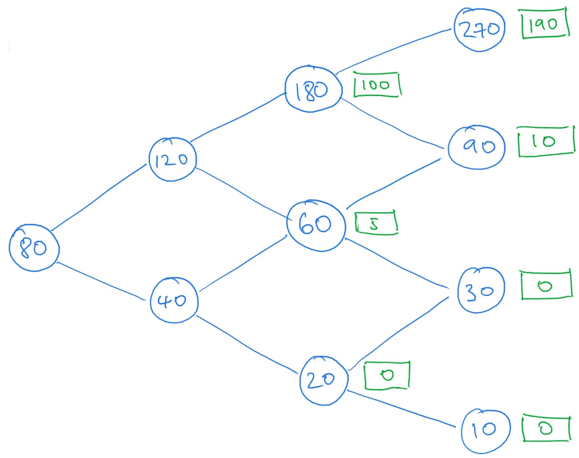
We now come to **STEP 3**, the key idea. Suppose we are sitting in one of the nodes at time $t = 2$, which we think of as the 'current' node. For example suppose we are at the uppermost node (labelled 180, the 'current' value of the stock). Looking forwards one step of time we can see that, if the stock price goes up our option is worth 190, whereas if the stock price goes down our option is worth 10. What we are seeing here is (an instance of) the one-period model! With contingent claim

$$\Phi(su) = 190, \quad \Phi(sd) = 10.$$

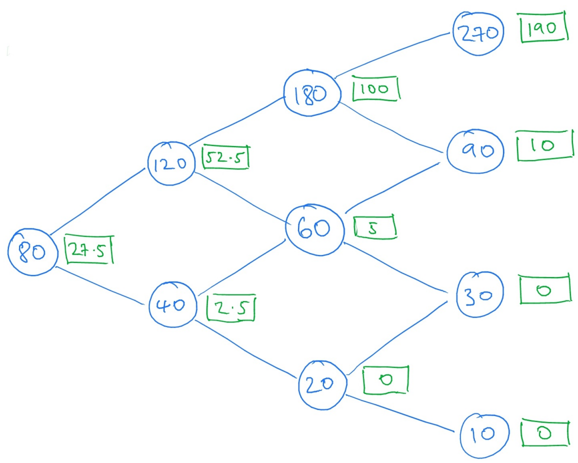
So, using the one-period risk-neutral valuation formula from Proposition 5.2.6 the value of our call option at our current node is

$$\frac{1}{1+0}(190 \cdot 0.5 + 10 \cdot 0.5) = 100.$$

We could apply the same logic to any of the nodes corresponding to time $t = 2$, and compute the value of our call option at that node:



If we now imagine ourselves sitting in one of the nodes at time $t = 1$, and look forwards one step in time, we again find ourselves faced with an instance of the one-period model. This allows us to compute the value of our call option at the $t = 1$ nodes; take for example the node labelled by 40 which, one step into the future, sees the contingent claim $\Phi(su) = 5, \Phi(sd) = 0$ and using (5.9) gives the value of the call option at this node as $\frac{1}{1+0}(h \cdot 0.5 + 0 \cdot 0.5) = 2.5$. Repeating the procedure on the other $t = 1$ node, and then also on the single $t = 0$ node gives us



Therefore, the value (i.e. the price) of our call option at time $t = 0$ is 27.5.

Although we have computed the price, we haven't yet computed a replicating portfolio, which is **STEP 4**. We could do it by solving lots of linear equations for our one-period models, as in Example 5.2.4, but since we have several steps a quicker way is to apply Proposition 5.2.6 and use the general formula we found in (5.8).

Starting at time $t = 0$, to replicate the contingent claim $\Phi(su) = 52.5$ and $\Phi(sd) = 2.5$ at time $t = 1$, equation (5.8) tells us that we want the portfolio

$$x_1 = \frac{1}{1+0}\frac{1.5 \cdot 2.5 - 0.5 \cdot 52.5}{1.5 - 0.5} = -22.5, \quad y_1 = \frac{1}{80}\frac{52.5 - 2.5}{1.5 - 0.5} = \frac{5}{8}.$$

The value of this portfolio at time 0 is

$$x_1 + 80y_1 = -22.5 + 80 \cdot \frac{5}{8} = 27.5$$

which is equal to the initial value of our call option.

We can then carry on forwards. For example, if the stock went up in between $t = 0$ and $t = 1$, then at time $t = 1$ we would be sitting in the node for $S_1 = 120$, labelled simply 120. Our portfolio (x_1, y_1) is now worth

$$x_1(1+0) + y_1 \cdot 120 = -22.5 + 120 \cdot \frac{5}{8} = 52.5,$$

equal to what is now the value of our call option. We use (5.8) again to calculate the portfolio we want to hold during time $1 \mapsto 2$, this time with $\Phi(su) = 100$ and $\Phi(sd) = 5$, giving $x_2 = -42.5$ and $y_2 = \frac{95}{120}$. You can check that the current value of the portfolio (x_2, y_2) is 52.5.

Next, suppose the stock price falls between $t = 1$ and $t = 2$, so our next node is $S_2 = 60$. Our portfolio (x_2, y_2) now becomes worth

$$x_2(1+0) + y_2 \cdot 60 = -42.5 + \frac{95}{120} \cdot 60 = 5,$$

again equal to the value our of call option. For the final step, we must replicate the contingent claim $\Phi(su) = 10$, $\Phi(sd) = 0$, which (5.8) tells us is done using $x_3 = -5$ and $y_3 = \frac{1}{6}$. Again, you can check that the value of this portfolio is 5.

Lastly, the stock price rises again to $S_3 = 90$. Our portfolio becomes worth

$$x_3(1+0) + y_3 \cdot 90 = -5 + \frac{1}{6} \cdot 90 = 10,$$

equal to the payoff from our call option.

To sum up, using (5.8) we can work out which portfolio we would want to hold, at each possible outcome of the stock changing value. At all times we would be holding a portfolio with current value equal to the current value of the call option. Therefore, this gives a self-financing portfolio strategy that replicates $\Phi(S_T)$.

5.7 Exercises

All questions use the notation u, d, p_u, p_d, s and r, which has been used throughout this chapter. In all questions we assume that the models are arbitrage free and complete: $d < 1 + r < u$.

On the one-period model

5.1 Suppose that we hold the portfolio $(1, 3)$ at time 0. What is the value of this portfolio at time 1?

5.2 Find portfolios that replicate the following contingent claims.

(a) $\Phi(S_1) = 1$

(b) $\Phi(S_1) = \begin{cases} 3 & \text{if } S_1 = su, \\ 1 & \text{if } S_1 = sd. \end{cases}$.

Hence, write down the values of these contingent claims at time 0.

5.3 Find the contingent claims $\Phi(S_1)$ for the following derivatives.

(a) A contract in which we promise to buy two units of stock at time $t = 1$ for strike price K.

(b) A European put option with strike price $K \in (sd, su)$ (see Section 5.3).

(c) A contract in which we promise that, if $S_1 = su$, we will sell one unit of stock at time $t = 1$ for strike price $K \in (sd, su)$ (and otherwise, if $S_1 = sd$ we do nothing).

(d) Holding both the contracts in (b) and (c) at once.

5.4 Let Π_t^{call} and Π_t^{put} be the price of European call and put options, both with the same strike price $K \in (sd, su)$, at times $t = 0, 1$.

(a) Write down formulae for Π_0^{call} and Π_0^{put}.

(b) Show that $\Pi_0^{call} - \Pi_0^{put} = s - \frac{K}{1+r}$.

On the binomial model

5.5 Write down the contingent claim of a European call option (that matures at time T).

5.6 Let $T = 2$ and let the initial value of a single unit of stock be $S_0 = 80$. Suppose that $p_u = p_d = 0.5$, that $u = 1.5$ and $d = 0.75$, and that $r = 0.1$. Draw out, in a tree-like diagram, the possible values of the stock price at times $t = 0, 1, 2$. Find the price, at time 0, of a European put option with strike price $K = 100$.

5.7 Recall that $(S_t)_{t=1}^T$ is the price of a single unit of stock. Find a condition on p_u, p_d, u, d that is equivalent to saying that S_t is a martingale under $\mathbb{P}$.

When is $M_t = \log S_t$ is a martingale under $\mathbb{P}$?

Challenge questions

5.8 Write a computer program (in a language of your choice) that carries out the pricing algorithm for the binomial model, for a general number n of time-steps.

Chapter 6

Convergence of random variables

A real number is a simple object; it takes a single value. As such, if a_n is a sequence of real numbers, $\lim_{n\to\infty} a_n = a$, means that the value of a_n converges to the value of a.

Random variables are more complicated objects. They take many different values, with different probabilities. Consequently, if $X_1, X_2, \ldots$ and X are random variables, there are many different ways in which we can try to make sense of the idea that $X_n \to X$. They are called *modes* of convergence, and are the focus of this chapter.

6.1 Modes of convergence

We say:

- $X_n \stackrel{\mathbb{P}}{\to} X$, known as convergence **in probability**, if given any $a > 0$,

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > a] = 0.$$

- $X_n \stackrel{a.s.}{\to} X$, known as **almost sure** convergence, if

$$\mathbb{P}\left[X_n \to X \text{ as } n \to \infty\right] = 1.$$

- $X_n \stackrel{L^p}{\to} X$, known as **convergence in** L^p, if

$$\mathbb{E}\left[|X_n - X|^p\right] \to 0 \text{ as } n \to \infty.$$

Here, $p \geq 1$ is a real number. We will be interested in the cases $p = 1$ and $p = 2$. The case $p = 2$ is sometimes known as convergence in **mean square**.

It is common for random variables to converge in some modes but not others, as the following example shows.

Example 6.1.1 *Let U be a uniform random variable on $[0, 1]$ and set*

$$X_n = n^2 \mathbb{1}\{U < 1/n\} = \begin{cases} n^2 & \textit{if } U < 1/n, \\ 0 & \textit{otherwise.} \end{cases}$$

Our candidate limit is $X = 0$, the random variable that takes the deterministic value 0.

If $X_m = 0$ for some $m \in \mathbb{N}$ then $X_n = 0$ for all $n \geq m$, which implies that $X_n \to 0$. So, we have

$$\mathbb{P}\left[\lim_{n\to\infty} X_n = 0\right] \geq \mathbb{P}[X_m = 0] = 1 - \frac{1}{m}.$$

Since this is true for any $m \in \mathbb{N}$, we have $\mathbb{P}[\lim_{n\to\infty} X_n = 0] = 1$, that is $X_n \stackrel{a.s.}{\to} 0$.

However, $\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = n^2 \frac{1}{n} = n$, which does not tend to 0 as $n \to \infty$. So X_n does not converge to 0 in L^1.

For any $0 < a \leq n^2$ we have $\mathbb{P}[|X_n - 0| > a] = \mathbb{P}[X_n > a] \leq \mathbb{P}[X_n = n^2] = \frac{1}{n}$, so as $n \to \infty$ we have $\mathbb{P}[|X_n - 0| > a] \to 0$, which means that we do have $X_n \stackrel{\mathbb{P}}{\to} 0$.

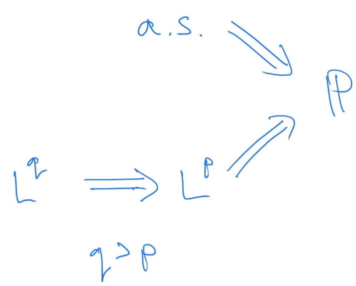
As we might hope, there are relationships between the different modes of convergence, which are useful to remember.

Lemma 6.1.2 *Let X_n, X be random variables.*

1. *If $X_n \stackrel{a.s.}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$.*
2. *If $X_n \stackrel{L^p}{\to} X$ then $X_n \stackrel{\mathbb{P}}{\to} X$.*
3. *Let $1 \leq p < q$. If $X_n \stackrel{L^q}{\to} X$ then $X_n \stackrel{L^p}{\to} X$.*

In all other cases, convergence in one mode does not imply convergence in another.

The proofs are not part of our course (they are part of MAS350/451/6051). We can summarise Lemma 6.1.2 with a diagram:



For convergence of real numbers, it was shown in MAS221 that if $a_n \to a$ and $a_n \to b$ then $a = b$, which is known as uniqueness of limits. For random variables, we have *almost sure uniqueness of limits*: if $X_n \stackrel{\mathbb{P}}{\to} X$ and $X_n \stackrel{\mathbb{P}}{\to} Y$ then $X = Y$ almost surely. Proving this fact is one of the challenge exercises, **6.5**. By Lemma 6.1.2, the same statement also holds if $\stackrel{\mathbb{P}}{\to}$ is replaced by $\stackrel{L^p}{\to}$ or $\stackrel{a.s.}{\to}$.

6.2 The dominated convergence theorem

A natural question to ask is, when does $\mathbb{E}[X_n] \to \mathbb{E}[X]$? We are interested (for use later on in the course) to ask when almost sure convergence implies that $\mathbb{E}[X_n] \to \mathbb{E}[X]$. As we can see from Example 6.1.1, in general it does not. We need an extra condition:

Theorem 6.2.1 (Dominated Convergence Theorem) *Let X_n, X be random variables such that:*

1. $X_n \stackrel{a.s.}{\to} X$.
2. *There exists a random variable $Y \in L^1$ such that, for all n, $|X_n| \le |Y|$ almost surely.*

Then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

The random variable Y is often known as the *dominating function* or *dominating random variable.*

Example 6.2.2 *Let X be a random variable in L^1 and set*

$$X_n = |X|\mathbb{1}_{\{|X|\ge n\}} = \begin{cases} X & \text{if } |X| \ge n, \\ 0 & \text{otherwise.} \end{cases}$$

We aim to use the dominated convergence theorem to show that $\mathbb{E}[X_n] \to 0$.

To check the first condition, since $X \in L^1$ we have $\mathbb{E}[|X|] < \infty$ and hence $X < \infty$ almost surely. If $n > |X|$ then $|X|\mathbb{1}_{\{|X|\ge n\}} = 0$, hence since $|X| < \infty$, as $n \to \infty$ we have $X_n = |X|\mathbb{1}_{\{|X|\ge n\}} \to 0$ almost surely; so we take $X = 0$.

To check the second condition, set $Y = |X|$ and then $\mathbb{E}[|Y|] = \mathbb{E}[|X|] < \infty$ so $Y \in L^1$. Also, $|X|\mathbb{1}_{\{X_n \ge n\}} \le |X| = Y$, so Y is a dominated random variable for (X_n). Hence, the dominated convergence theorem applies and $\mathbb{E}[X_n] \to \mathbb{E}[X] = 0$.

If our sequence of random variables (X_n) has $|X_n| \le c$ for some deterministic constant c, then the dominating function can be taken to be (the deterministic random variable) c. This case is the a very common application.

The dominated convergence theorem is one of the most important theorems in modern probability theory. Our focus, however, is on stochastic processes rather than theory. As such, although we will make use of the dominated convergence theorem in some of our proofs, we will not fully appreciate the extent of its importance to probability – a taste of this can be found in MAS350/451/6051, along with its proof.

Remark 6.2.3 ($\star$) *The dominated convergence theorem holds for conditional expectation too; that is with $\mathbb{E}[\cdot]$ replaced by $\mathbb{E}[\cdot \,|\, \mathcal{G}]$. We won't need this result as part of our course.*

6.3 Exercises

On convergence of random variables

6.1 Let (X_n) be a sequence of independent random variables such that

$$X_n = \begin{cases} 2^{-n} & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases}$$

Show that $X_n \to 0$ in L^1 and almost surely. Deduce that also $X_n \to 0$ in probability.

6.2 Let X_n, X be random variables.

(a) Suppose that $X_n \xrightarrow{L^1} X$ as $n \to \infty$. Show that $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

(b) Give an example where $\mathbb{E}[X_n] \to \mathbb{E}[X]$ but X_n does not converge to X in L^1.

6.3 Let U be a random variable that takes values in $(1, \infty)$. Define $X_n = U^{-n}$. Use the dominated convergence theorem to show that $\mathbb{E}[X_n] \to 0$.

6.4 Let (X_n) be the sequence of random variables from **6.1**. Define $Y_n = X_1 + X_2 + \ldots + X_n$.

(a) Show that, for all $\omega \in \Omega$, the sequence $Y_n(\omega)$ is increasing and bounded.

(b) Deduce that the limit $Y = \lim_{n\to\infty} Y_n$ exists almost surely.

(c) Write down the distribution of Y_1, Y_2 and Y_3.

(d) Suggest why we might guess that Y has a uniform distribution on $[0, 1]$.

(e) Prove that Y_n has a uniform distribution on $\{k2^{-n}\,;\, k = 0, 1, \ldots, 2^n - 1\}$.

(f) Prove that Y has a uniform distribution on $[0, 1]$.

Challenge questions

6.5 Show that if $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{\mathbb{P}} Y$ then $X = Y$ almost surely.

6.6 Let (X_n) be a sequence of independent random variables such that $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = 0] = \frac{1}{2}$. Show that (X_n) does not converge in probability and deduce that (X_n) also does not converge in L^1, or almost surely.

Chapter 7

Stochastic processes and martingale theory

In this section we study two important results from the theory of martingales and introduce the concept of a stopping time. We use our results to analyse the behaviour of stochastic processes.

From now on we write

$$\min(a,b) = a \wedge b, \qquad \max(a,b) = a \vee b.$$

This is common notation in the field of stochastic processes.

7.1 The martingale transform

Recall Definition 5.5.1: A stochastic process $C = (C_n)_{n=1}^{\infty}$ is said to be *previsible* if, for all n, C_n is $\mathcal{F}_{n-1}$ measurable.

The difference is that, in an adapted process (X_n), the value of X becomes fully known to us, using information from $\mathcal{F}_n$, at time n. In a previsible process (C_n), the value of C_n becomes known at time $n-1$.

If M is a stochastic process and C is previsible process, we define the *martingale transform* of S by M

$$(C \circ M)_n = \sum_{i=1}^{n} C_i(M_i - M_{i-1}).$$

Here, by convention, we set $(C \circ M)_0 = 0$.

If M is a martingale, the process $(C \circ M)_n$ can be thought of as our winnings after n plays of a game. Here, at round i, a bet of C_i is made, and the change to our resulting wealth is $C_i(M_i - M_{i-1})$. For example, if $C_i \equiv 1$ and M_n is the simple random walk $M_n = \sum_1^n X_i$ then $M_i - M_{i-1} = X_{i-1}$, so we win/lose each round with even chances; we bet 1 on each round, if we win we get our money back doubled, if we lose we get nothing back.

Theorem 7.1.1 *If M is a martingale and C is previsible and bounded, then $(C \circ M)_n$ is also a martingale.*

Similarly, if M is a supermartingale (resp. submartingale), and C is previsible, bounded and non-negative, then $(C \circ M)_n$ is also a supermartingale martingale (resp. submartingale).

PROOF: Let M be a martingale. Write $Y = C \circ M$. We have $C_n \in \mathcal{F}_{n-1}$ and $X_n \in \mathcal{F}_n$, so Proposition 2.2.4 implies that $Y_n \in m\mathcal{F}_n$. Since $|C| \leq c$ for some c, we have

$$\mathbb{E}|Y_n| \leq \sum_{k=1}^{n} \mathbb{E}|C_k(M_k - M_{k-1})| \leq c \sum_{k=1}^{n} \mathbb{E}|M_k| + \mathbb{E}|M_{k-1}| < \infty.$$

So $Y_n \in L^1$. Since C_n is $\mathcal{F}_{n-1}$-measurable, by linearity of conditional expectation, the taking out what is known rule and the martingale property of M, we have

$$\begin{aligned}
\mathbb{E}[Y_n \,|\, \mathcal{F}_{n-1}] &= \mathbb{E}[Y_{n-1} + C_n(M_n - M_{n-1}) \,|\, \mathcal{F}_{n-1}] \\
&= Y_{n-1} + C_n \mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \\
&= Y_{n-1} + C_n(\mathbb{E}[M_n|\mathcal{F}_{n-1}] - M_{n-1}) \\
&= Y_{n-1}.
\end{aligned}$$

Hence Y is a martingale.

The argument is easily adapted to prove the second statement, e.g. for a supermartingale M, $\mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \leq 0$. Note that in these cases it is important that C is non-negative. ■

7.2 Roulette

The martingale transform is a useful theoretical tool, but it also provides a framework to model casino games. We illustrate this with Roulette.

In roulette, a metal ball lies inside of a spinning wheel. The wheel is divided into 37 segments, of which 18 are black, 18 are red, and 1 is green. The wheel is spun, and the ball spins with it, eventually coming to rest in one of the 37 segments. If the roulette wheel is manufactured properly, the ball lands in each segment with probability $\frac{1}{37}$ and the result of each spin is independent.

On each spin, a player can bet an amount of money C. The player chooses either red or black. If the ball lands on the colour of their choice, they get their bet of C returned and win an additional C. Otherwise, the casino takes the money and the player gets nothing.

The key point is that players can only bet on red or black. If the ball lands on green, the casino takes *everyones* money.

Remark 7.2.1 *Thinking more generally, most casino games fit into this mould – there is a very small bias towards the casino earning money. This bias is known as the 'house advantage'.*

In each round of roulette, a players probability of winning is $\frac{18}{37}$ (it does not matter which colour they pick). Let (X_n) be a sequence of i.i.d. random variables such that

$$X_n = \begin{cases} 1 & \text{with probability } \frac{18}{37} \\ -1 & \text{with probability } \frac{19}{37} \end{cases}$$

Naturally, the first case corresponds to the player winning game n and the second to losing. We define

$$M_n = \sum_{i=1}^{n} X_n.$$

Then, the value of $M_n - M_{n-1} = X_n$ is 1 if the player wins game n and -1 if they lose. We take our filtration to be generated by (M_n), so $\mathcal{F}_n = \sigma(M_i\,;\, i \le n)$.

A player cannot see into the future. So the sequence of bets (C_n) that they place must be a previsible process. The total profit/loss of the player over time is the martingale transform

$$(C \circ M)_n = \sum_{i=1}^{n} C_n(M_n - M_{n-1}).$$

We'll now show that (M_n) is a supermartingale.We have $M_n \in m\mathcal{F}_n$ and since $|M_n| \le n$ we also have $M_n \in L^2$. Lastly,

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \mathbb{E}\left[X_{n+1} + M_n \,|\, \mathcal{F} - n\right] \\
&= \mathbb{E}[X_{n+1}|\mathcal{F}_n] + M_n \\
&= \mathbb{E}[X_{n+1}] + M_n \\
&\le M_n.
\end{aligned}$$

Here, the second line follows by linearity and the taking out what is known rule. The third line follows because X_{n+1} is independent of $\mathcal{F}_n$, and the last line follows because $\mathbb{E}[X_{n+1}] = \frac{-1}{37} < 0$.

So, (M_n) is a supermartingale and (C_n) is previsible. Theorem 7.1.1 applies and tells us that $(C \circ M)_n$ is a supermartingale.

7.3 The optional stopping theorem

Previsibility captures the idea that we cannot see into the future. This leaves us with the question, as time passes, in what circumstances can we tell if an event has already occurred?

Definition 7.3.1 *A map $T : \Omega \to \{0, 1, 2, \dots, \infty\}$ is called a $(\mathcal{F}_n)$ stopping time if, for all n, $\{T = n\}$ is $\mathcal{F}_n$ measurable.*

Equivalently, we could say that T is a stopping time if $\{T \le n\}$ is $\mathcal{F}_n$ measurable for all n. To see why, recall the definition of a σ-algebra and note that

$$\{T \le n\} = \bigcup_{i \le n} \{T = i\}, \quad \{T = n\} = \{T \le n\} \setminus \{T \le n-1\}.$$

A stopping time is a random time with the property that, if we have only information from $\mathcal{F}_n$ accessible to us at time n, we are able to decide at any n whether or not T has already happened.

The simplest example of a stopping time is a constant random variable; if $t \in \mathbb{R}$ and $T = t$ then $\{T = n\} = \{t = n\}$, which is empty if $t \ne n$ and equal to Ω if $t = n$. However, in general stopping times are useful because they allow us to describe the random behaviour of stochastic processes.

Example 7.3.2 *Let $S_n = \sum_{i=1}^n X_i$ be the simple symmetric random walk, with $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$. Let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then, for any $a \in \mathbb{N}$, the time*

$$T = \inf\{n \ge 0\,;\, S_n = a\},$$

which is the first time S_n takes the value a, is a stopping time. It is commonly called the hitting time of a. To see that T is a stopping time we note that

$$\begin{aligned}
\{T = n\} &= \{S_n = a\} \cap \{S_i \ne a \textit{ for all } i < a\} \\
&= \{S_n = a\} \cap \left(\bigcap_{i=0}^{n-1} \Omega \setminus \{S_n = a\} \right) \\
&= S_n^{-1}(a) \cap \left(\bigcap_{i=0}^{n-1} \Omega \setminus S_n^{-1}(a) \right).
\end{aligned}$$

Since S_n is $\mathcal{F}_n$ measurable (by Proposition 2.2.4), the above equation shows that $\{T = n\} \in \mathcal{F}_n$.

Lemma 7.3.3 *Let S and T be stopping times with respect to the filtration $(\mathcal{F}_n)$. Then $S \wedge T$ is also a $(\mathcal{F}_n)$ stopping time.*

PROOF: Note that

$$\{S \wedge T \le n\} = \{S \le n\} \cup \{T \le n\}.$$

Since S and T are stopping times, both the sets on the right hand side of the above are events in the σ-field $\mathcal{F}_n$. Hence, the event on the left hand side is also in $\mathcal{F}_n$. ■

If T is a stopping time and M is a stochastic process, we define M^T to be the process

$$M_n^T = M_{n \wedge T}.$$

Here $a \wedge b$ denotes the minimum of a and b. To be precise, this means that $M_n^T(\omega) = M_{n \wedge T(\omega)}(\omega)$ for all $\omega \in \Omega$. In Example 7.3.2, S^T would be the random walk S which is stopped (i.e. never moves again) when (if!) it reaches state a.

Lemma 7.3.4 *Let M_n be a martingale (resp. supmartingale, supermartingale) and let T be a stopping time. Then M^T is also a martingale (resp. supmartingale, supermartingale).*

PROOF: Let $C_n := \mathbb{1}\{T \geq n\}$. Note that $\{T \geq n\} = \Omega \setminus \{T \leq n-1\}$, so $\{T \geq n\} \in \mathcal{F}_{n-1}$. By Lemma 2.4.2 $C_n \in m\mathcal{F}_{n-1}$. That is, (C_n) is a previsible process. Moreover,

$$(C \circ M)_n = \sum_{k=1}^{n} \mathbb{1}_{k \leq T}(M_k - M_{k-1}) = \sum_{k=1}^{n \wedge T}(M_k - M_{k-1}) = M_{T \wedge n} - M_0.$$

The last equality holds because the sum is telescoping (the middle terms all cancel each other out). Hence, if M is a martingale (resp. submartingale, supermartingale), $C \circ M$ is also a martingale (resp. submartingale, supermartingale) by Theorem 7.1.1. ■

Theorem 7.3.5 (Doob's Optional Stopping Theorem) *Let M be martingale (resp. submartingale, supermartingale) and let T be a stopping time. Then*

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]$$

(resp. $\geq$, $\leq$) if any one of the following conditions hold:

a. T is bounded.

b. M is bounded and $\mathbb{P}[T < \infty] = 1$.

c. $\mathbb{E}[T] < \infty$ and there exists $c \in \mathbb{R}$ such that $|M_n - M_{n-1}| \leq c$ for all n.

PROOF: We'll prove this for the supermartingale case. The submartingale case then follows by considering $-M$, and the martingale case follows since martingales are both supermartingales and submartingales.

Note that

$$\mathbb{E}[M_{n \wedge T} - M_0] \leq 0, \tag{7.1}$$

because M^T is a supermartingale, by Lemma 7.3.4. For (a), we take $n = \sup_\omega T(\omega)$ and the conclusion follows.

For (b), we use the dominated convergence theorem to let $n \to \infty$ in (7.1). As $n \to \infty$, almost surely $n \wedge T(\omega)$ is eventually equal to $T(\omega)$ (because $\mathbb{P}[T < \infty] = 1$), so $M_{n \wedge T} \to M_T$ almost surely. Since M is bounded, $M_{n \wedge T}$ and M_T are also bounded. So $\mathbb{E}[M_{n \wedge T}] \to \mathbb{E}[M_T]$ and taking limits in (7.1) obtains $\mathbb{E}[M_T - M_0] \leq 0$, which in turn implies that $\mathbb{E}[M_T] \leq \mathbb{E}[M_0]$.

For (c), we will also use the dominated convergence theorem to let $n \to \infty$ in (7.1), but now we need a different way to check its conditions. We observe that

$$|M_{n \wedge T} - M_0| = \left| \sum_{k=1}^{n \wedge T}(M_k - M_{k-1}) \right| \leq T \sup_{n \in \mathbb{N}} |M_n - M_{n-1}|.$$

Since $\mathbb{E}[T(\sup_n |M_n - M_{n-1}|)] \leq c\mathbb{E}[T] < \infty$, we can use the Dominated Convergence Theorem to let $n \to \infty$, and the results follows as in (b). ■

7.4 Hitting times of random walks

We can now use the optional stopping theorem to tell us about hitting probabilities and expected hitting times of various stochastic processes. In this section we focus on random walks. Two examples with urn processes can be found in the exercises.

7.4.1 Asymmetric random walk

We recall the asymmetric random walk from Section 4.1.2. Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables. Let $p+q=1$ with $p,q\in[0,1]$, $p\neq q$ and suppose that

$$\mathbb{P}[X_i=1]=p,\quad \mathbb{P}[X_i=-1]=q.$$

The asymmetric random walk is the stochastic process

$$S_n=\sum_{i=1}^{n}X_i.$$

Set $\mathcal{F}_n=\sigma(X_1,\ldots,X_n)$ and note that $(\mathcal{F}_n)$ is the natural filtration of S_n. Recall that we showed in Section 4.1.2 that the stochastic process $S_n-n(p-q)$ was a martingale.

There is another martingale associated to the asymmetric random walk. Define

$$M_n=(q/p)^{S_n}.$$

We will now show that M_n is a martingale.

Since $X_i\in m\mathcal{F}_n$ for all $i\leq n$, by Proposition 2.2.4 we have $M_n\in m\mathcal{F}_n$. We have $|X_i|\leq 1$ so $|M_n|\leq (q/p)^n$, which implies that $S_n\in L^1$ for all n. Moreover,

$$\begin{aligned}\mathbb{E}[M_{n+1}|\mathcal{F}_n]&=(q/p)^{S_n}\mathbb{E}[(q/p)^{X_{n+1}}|\mathcal{F}_n]\\&=(q/p)^{S_n}\mathbb{E}[(q/p)^{X_{n+1}}]\\&=(q/p)^{S_n}=M_n.\end{aligned}$$

Here we use the taking out what is known rule, followed by the fact that X_{n+1} is independent of $\mathcal{F}_n$ and the relationship between conditional expectation and independence. To deduce the final step we use that $\mathbb{E}[(q/p)^{X_{n+1}}]=p(q/p)^1+q(q/p)^{-1}=p+q=1$.

Our next plan is to use the optional stopping theorem, applied to the martingale M_n, to obtain information about the hitting times of the asymmetric random walk. Let $T_a=\inf\{n:S_n=a\}$ and $T=T_a\wedge T_b$ for integer $a<0<b$. We aim to calculate $\mathbb{P}[T=T_a]$ and $\mathbb{P}[T=T_b]$. We can show that T_a is a stopping time by noting that

$$\{T_a\leq n\}=\bigcup_{i=0}^{n}\{S_n\leq a\}.$$

Similarly, T_b is a stopping time and it follows from Lemma 7.3.3 that T is also a stopping time.

We now look to apply the optional stopping theorem, using the (b) conditions. For this, we'll need the following lemma.

Lemma 7.4.1 *It holds that* $\mathbb{E}[T]<\infty$.

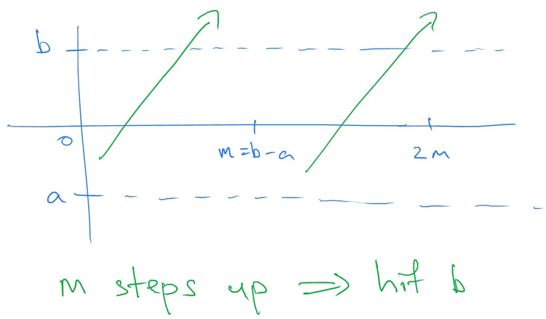
PROOF: Let $m = b - a$. Divide up the random variables (X_n) into sets $A_1, A_2, \ldots$ as follows:

$$\overbrace{X_1, X_2, X_3, \ldots, X_m}^{A_1}, \overbrace{X_{m+1}, X_{m+2}, \ldots, X_{2m}}^{A_2}, \overbrace{X_{2m+1}, X_{2m+2}, \ldots, X_{3m}}^{A_3}, \overbrace{X_{3m+1}, \ldots}^{\ldots}. \tag{7.2}$$

So as each A_k contains precisely m of the X_is.

Let $E_k = \{\text{for all } X \in A_k,\ X = 1\}$ be the event that all random variables in A_k are equal to one. Note that if the event E_k occurs then our random walk moves up by m steps, during time $(k-1)m, \ldots, km$.

If T has not happened before time $(k-1)m$ then $a < S_{(k-1)m} < b$. Then, if the event E_k then occurs, we will have $S_{km} \geq a$ and hence $T \leq km$. This is best illustrated with a picture:



We can think of the sequence of events $E_1, E_2, \ldots$ as one chance after another that our random walker has to exit $[a, b]$.

By independence, $\mathbb{P}[E_k] = p^m$. Hence, the random variable

$$K = \min\{k \in \mathbb{N};\ E_k \text{ occurs}\}$$

is a geometric random variable with success parameter p^m. This means that $K < \infty$ almost surely and that $\mathbb{E}[K] = p^{-m} < \infty$. By definition of K, the event E_K occurs so $T \leq Km$ and by monotonicity of $\mathbb{E}$ we have $\mathbb{E}[T] \leq m\mathbb{E}[K] < \infty$. ■

From above, we have that M is a martingale. Hence by Lemma 7.3.4, M^T is also a martingale. By definition of T,

$$(q/p)^a \leq M_n^T \leq (q/p)^b$$

for all n, hence M^T is a bounded martingale. Lemma 7.4.1 implies that $\mathbb{P}[T < \infty] = 1$, so we have that condition (b) of the optional stopping theorem holds for the martingale M^T and the stopping time T. Therefore,

$$\mathbb{E}[M_T^T] = \mathbb{E}[M_0^T]$$

but $M_T^T = M_{T \wedge T} = M_T$ and $M_0^T = M_{0 \wedge T} = M_0 = 1$. So we have

$$\mathbb{E}[M_T] = 1.$$

Our next aim is to calculate the probabilities $\mathbb{P}[T = T_a]$ and $\mathbb{P}[T = T_b]$. That is, we want to know which of the two boundaries a and b we actually hit at time T (for example, $\{T = T_a\} = \{S_T = a\}$ is the event that we hit a at time T).

Since $\mathbb{P}[T < \infty] = 1$, we must hit one or other boundary, so we have that

$$\mathbb{P}[T = T_a] + \mathbb{P}[T = T_b] = 1.$$

By partitioning the expectation $\mathbb{E}[M_T]$ on whether or not $\{T = T_a\}$, we have

$$\begin{aligned}
1 &= \mathbb{E}[M_T] \\
&= \mathbb{E}[M_T \mathbb{1}\{T = T_a\}] + \mathbb{E}[M_T \mathbb{1}\{T = T_b\}] \\
&= \mathbb{E}\left[\left(\frac{q}{p}\right)^a \mathbb{1}\{T = T_a\}\right] + \mathbb{E}\left[\left(\frac{q}{p}\right)^b \mathbb{1}\{T = T_b\}\right] \\
&= \mathbb{P}[T = T_a]\left(\frac{q}{p}\right)^a + \mathbb{P}[T = T_b]\left(\frac{q}{p}\right)^b.
\end{aligned}$$

Solving these two linear equations (recall that $p \neq q$) gives that

$$\mathbb{P}[T = T_a] = \frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a}. \tag{7.3}$$

and therefore also

$$\mathbb{P}[T_b = T] = 1 - \mathbb{P}[T = T_a] = \frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}.$$

7.4.2 Symmetric random walk

We now recall the symmetric random walk from Section 4.1.1. Our plan is much the same; we aim to use martingales and optional stopping to investigate the hitting times of the symmetric random walk. This case is a bit harder than the asymmetric one.

Let $(X_i)_{i=1}^{\infty}$ be a sequence of i.i.d. random variables where

$$\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}.$$

The symmetric random walk is the stochastic process

$$S_n = \sum_{i=1}^{n} X_i.$$

Set $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and note that $(\mathcal{F}_n)$ is a filtration. We have already seen that S_n is a martingale.

Like the asymmetric case, there is another martingale associated to S_n. In fact, there is a whole family of them. Let $\theta \in \mathbb{R}$ and define

$$M_n^{(\theta)} = \frac{e^{\theta S_n}}{(\cosh\theta)^n}.$$

Note that $M_n^{(\theta)} = \prod_{i=1}^{n}(e^{\theta X_i}/\cosh\theta)$. Since $X_i \in m\mathcal{F}_n$ for all $i \leq n$, $S_n \in m\mathcal{F}_n$ for all n by Proposition 2.2.4. Since $|S_n| \leq n$ we have $|M_n| \leq \frac{e^{\theta n}}{(\cosh\theta)^n} < \infty$, hence $M_n \in L^1$. We have also that

$$\begin{aligned}
\mathbb{E}[M_{n+1}|\mathcal{F}_n] &= \left(\prod_{i=1}^{n} \frac{e^{\theta X_i}}{\cosh\theta}\right) \mathbb{E}\left[\frac{e^{\theta X_{n+1}}}{\cosh\theta}\middle|\mathcal{F}_n\right] \\
&= M_n \mathbb{E}\left[\frac{e^{\theta X_{n+1}}}{\cosh\theta}\right] \\
&= M_n.
\end{aligned}$$

Here we use the taking out what is known rule, the fact that X_{n+1} is independent of $\mathcal{F}_n$ and the relationship between conditional expectation and independence. To deduce the final line we note that $\mathbb{E}[e^{\theta X_i}/\cosh\theta] = \frac{1}{2}(e^{\theta} + e^{-\theta})/\cosh\theta = 1$.

Let

$$T = \inf\{n\,;\, S_n = 1\},$$

which we have seen is a stopping time in Example 7.3.2. It is not obvious whether $\mathbb{P}[T = \infty]$ is equal to or greater than zero, but with the help of M_n and the optional stopping theorem we can show:

Lemma 7.4.2 *It holds that $\mathbb{P}[T < \infty] = 1$.*

PROOF: Let us consider $\theta > 0$. By Lemma 7.3.3, $T \wedge n$ is a stopping time, and since $T \wedge n \leq n$ it is a bounded stopping time. We therefore have condition (a) of the optional stopping theorem and can apply it to deduce that

$$\mathbb{E}[M_0] = 1 = \mathbb{E}\left[M_{T\wedge n}^{(\theta)}\right].$$

We now apply the dominated convergence theorem to let $n \to \infty$ in the rightmost term. To do so we make two observations: (1) $M_{T\wedge n} = e^{\theta S_{T\wedge n}}/(\cosh\theta)^n$ is bounded above by e^θ since $\cosh\theta \geq 1$ and $S_{T\wedge n} \in (-\infty, 1]$; (2) as $n \to \infty$, $M^{(\theta)}_{T\wedge n} \to M^{(\theta)}_T$, where the latter is defined to be 0 if $T = \infty$. Note that (2) uses that $\theta > 0$. So the dominated convergence theorem gives us

$$1 = \mathbb{E}\left[M^{(\theta)}_T\right].$$

Noting that $M^{(\theta)}_T = e^{\theta S_T}/(\cosh\theta)^T$ and $S_T = 1$ when $T < \infty$, we thus have

$$\mathbb{E}\left[\frac{e^\theta}{(\cosh\theta)^{-T}}\right] = 1 \tag{7.4}$$

for $\theta > 0$ (where we allow $\frac{e^\theta}{\infty} = 0$ for the case $T = \infty$).

If $T = \infty$, then $(\cosh\theta)^{-T} = 0$ for all $\theta \neq 0$. If $T < \infty$, then $(\cosh\theta)^{-T} \to 1$ as $\theta \to 0$. Noting that $0 \leq (\cosh\theta)^{-T} \leq 1$, we can apply the dominated convergence theorem and let $\theta \to 0$, with the result that

$$\mathbb{E}[\mathbb{1}\{T < \infty\}] = 1.$$

Hence $\mathbb{P}[T < \infty] = 1$. ■

Even though we now know that $\mathbb{P}[T < \infty] = 1$, we are not able to apply optional stopping to (S_n) and T; because (S_n) is unbounded and we do not know if $\mathbb{E}[T] < \infty$. In fact, in this case $\mathbb{E}[T] = \infty$ and the optional stopping theorem does not apply. Instead, we can use the optional stopping theorem to *deduce* that $\mathbb{E}[T] = \infty$, as follows.

Suppose, for a contradiction, that $\mathbb{E}[T] < \infty$. Then we could apply the optional stopping theorem to (S_n) and T using the condition (c). Hence $\mathbb{E}[S_0] = 0 = \mathbb{E}[S_T]$. But, $\mathbb{P}[T < \infty] = 1$ which means, by definition of T, that $S_T = 1$ almost surely. This is a contradiction so we must have $\mathbb{E}[T] = \infty$.

7.5 Exercises

On the martingale transform

7.1 Let $(\mathcal{F}_n)$ be a filtration and let S, T be stopping times such that $S \leq T$ almost surely. Define

$$A_n = \begin{cases} 1 & \text{if } S < n \leq T \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that the stochastic process (A_n) is previsible.

(b) Let (X_n) be a stochastic process. Show that $(A \circ X)_n = X_{T \wedge n} - X_{S \wedge n}$.

(c) If (X_n) is a supermartingale with $X_0 = 0$, show that $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_{S\wedge n}]$.

On optional stopping

7.2 Let (S_n) be the asymmetric random walk, started from $S_0 = 0$. Let $a < 0 < b$ be natural numbers and define the stopping times $T_a = \inf\{n\,;\, S_n = a\}$, $T_b = \inf\{n\,;\, S_n = b\}$ and $T = T_a \wedge T_b$.

(a) Show that $\mathbb{E}[S_T] = (p - q)\mathbb{E}[T]$.

(b) Calculate $\mathbb{E}[S_T]$ directly using equation (7.3) and hence calculate $\mathbb{E}[T]$.

7.3 Let S and T be stopping times with respect to the filtration $\mathcal{F}_n$.

(a) Show that $\max(S, T)$ is a stopping time.

(b) Suppose $S \leq T$ almost surely. Is it necessarily true that $T - S$ is a stopping time?

7.4 Recall the urn process from Section 4.2. Recall that B_n is the number of black balls added to the urn at or before time n, and that $M_n = \frac{B_n+1}{n+2}$ is the fraction of black balls in the urn at time n. Let T be the time at which the first black ball is drawn from the urn.

(a) Show that $\mathbb{P}[T \geq n] = \frac{1}{n-1}$.

(b) Use the optional stopping theorem to show that $\mathbb{E}[M_T] = \frac{1}{2}$ and $\mathbb{E}[\frac{1}{T+2}] = \frac{1}{4}$.

7.5 Let $m \in \mathbb{N}$ and $m \geq 2$. At time $n = 0$, an urn contains $2m$ balls, of which m are red and m are blue. At each time $n = 1, \dots, 2m$ we draw a single ball from the urn; we do not replace it. Therefore, at time n the urn contains $2m - n$ balls.

Let N_n denote the number of red balls remaining in the urn at time n. For $n = 0, \dots, 2m-1$ let

$$P_n = \frac{N_n}{2m - n}$$

be the fraction of red balls remaining after time n.

(a) Show that P_n is a martingale, with respect to a natural filtration that you should specify.

(b) **[Challenge question]** Let T be the first time at which the ball that we draw is red. Note that $T < 2m$, because the urn initially contains at least 2 red balls. Show that the probability that the $(T + 1)^{st}$ ball is red is $\frac{1}{2}$.

Chapter 8

Further theory of stochastic processes (Δ)

In this chapter we develop some further tools for analysing stochastic processes: the Borel-Cantelli lemmas and the martingale convergence theorem. We are primarily interested in the behaviour of stochastic processes as time tends to infinity. As our main example, in Section 4.3 we will look at the Galton-Watson branching process.

Note that this whole chapter is marked with a (Δ) – it is for independent study in MAS452/6052 but it is not part of MAS352.

8.1 The Borel-Cantelli lemmas (Δ)

The Borel-Cantelli lemmas are a tool for understanding the tail behaviour of a sequence (E_n) of events. The key definitions are

$$\begin{aligned} \{E_n \text{ i.o.}\} = \{E_n, \text{ infinitely often}\} &= \bigcap_m \bigcup_{n\geq m} E_n = \{\omega : \omega \in E_n \text{ for infinitely many } n\} \\ \{E_n \text{ e.v.}\} = \{E_n, \text{ eventually}\} &= \bigcup_m \bigcap_{n\geq m} E_n = \{\omega : \omega \in E_n \text{ for all sufficiently large } n\}. \end{aligned}$$

The set $\{E_n \text{ i.o.}\}$ is the the event that infinite many of the events E_n occur. The set $\{E_n \text{ e.v.}\}$ is the event that, for some (random) N, all the events E_n for which $n \geq N$ occur. We include the set theoretic definitions for completeness; for our purposes the verbal definition is sufficient.

For example, we might take an infinite sequence of coin tosses and choose E_n to be the event that the n^{th} toss is a head. Then $\{E_n \text{ i.o.}\}$ is the event that infinitely many heads occur, and $\{E_n \text{ e.v.}\}$ is the event that, after some point, all remaining tosses show heads.

Note that by straightforward set algebra,

$$\Omega \setminus \{E_n \text{ i.o.}\} = \{\Omega \setminus E_n \text{ e.v.}\}. \tag{8.1}$$

In our coin tossing example, $\Omega \setminus E_n$ is the event that the n^{th} toss is a tail. So (8.1) says that 'there are not infinitely many heads' if and only if 'eventually, we see only tails'.

The Borel-Cantelli lemmas, respectively, give conditions under which the probability of $\{E_n \text{ i.o.}\}$ is either 0 or 1. For completeness we will include the proofs but they are outside of the scope of our course (they are part of MAS350/451/6051). To be precise:

Lemma 8.1.1 (First Borel-Cantelli Lemma) *Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of events and suppose $\sum_{n=1}^\infty \mathbb{P}[E_n] < \infty$. Then $\mathbb{P}[E_n \text{ i.o.}] = 0$.*

PROOF: $(\star)$ Since $\sum_{n=1}^\infty \mathbb{P}[E_n] < \infty$, we have $\lim_{N\to\infty} \sum_{n=N}^\infty \mathbb{P}[E_n] = 0$. Hence, we have $\inf_N \sum_{n=N}^\infty \mathbb{P}[E_N] = 0$. Therefore

$$\mathbb{P}\left[\bigcap_N \bigcup_{n\geq N} E_n\right] \leq \inf_N \mathbb{P}\left[\bigcup_{n\geq N} E_N\right] \leq \inf_N \sum_{n=N}^\infty P[E_n] = 0,$$

which completes the proof. ■

For example, suppose that (X_n) are random variables that take the values 0 and 1, and that $\mathbb{P}[X_n = 1] = \frac{1}{n^2}$ for all n. Then $\sum_n \mathbb{P}[X_n = 1] = \sum_n \frac{1}{n^2} < \infty$ so, by Lemma 8.1.1, $\mathbb{P}[X_n = 1 \text{ i.o.}] = 0$, which by (8.1) means that $\mathbb{P}[X_n = 0 \text{ e.v.}] = 1$. In turn, this means that $X_n \to 0$ almost surely. Note that we did not require the (X_n) to be independent!

Lemma 8.1.2 (Second Borel-Cantelli Lemma) *Let $(E_n)_{n\in\mathbb{N}}$ be a sequence of independent events and suppose that $\sum_{n=1}^\infty \mathbb{P}[E_n] = \infty$. Then $\mathbb{P}[E_n \text{ i.o.}] = 1$.*

PROOF: $(\star)$ Write $E_n^c = \Omega \setminus E_n$. We will show that $\mathbb{P}[E_n^c \text{ e.v.}] = 0$, which by (8.1) implies the stated result. Note that

$$\mathbb{P}[E_n^c \text{ e.v.}] \leq \sum_{N=1}^\infty \mathbb{P}\left[\bigcap_{n\geq N} E_n^c\right]. \tag{8.2}$$

Moreover, since the (E_n) are independent, so are the (E_n^c), so

$$\mathbb{P}\left[\bigcap_{n\geq N} E_n^c\right] = \prod_{n=N}^\infty \mathbb{P}[E_n^c] = \prod_{n=N}^\infty (1 - \mathbb{P}[E_n]) \leq \prod_{n=N}^\infty e^{-\mathbb{P}[E_n]} = \exp\left(-\sum_{n=N}^\infty \mathbb{P}[E_n]\right) = 0.$$

Here we use that $1 - x \leq e^{-x}$ for $x \in [0,1]$. ■

For example, suppose that (X_n) are i.i.d. random variables such that $\mathbb{P}[X_n = 1] = \frac{1}{2}$ and $\mathbb{P}[X_n = -1] = \frac{1}{2}$. Then $\sum_n \mathbb{P}[X_n = 1] = \infty$ and, by Lemma 8.1.2, $\mathbb{P}[X_n = 1 \text{ i.o.}] = 1$. By symmetry, we have also $\mathbb{P}[X_n = 0 \text{ i.o.}] = 1$, so we can conclude that (X_n) is not almost surely convergent.

Note that, in both our examples above, we used the Borel-Cantelli lemmas to deduce a statement about almost sure convergence. The Borel-Cantelli lemmas are often useful for proving, or dis-proving, almost sure convergence.

Since both the Borel-Cantelli lemmas come down to summing a series, a useful fact to remember from real analysis is that, for $p \in \mathbb{R}$,

$$\sum_{n=1}^\infty n^{-p} < \infty \quad \Leftrightarrow \quad p > 1.$$

This follows from **1.7** and the integral test for convergence of series.

8.2 The martingale convergence theorem (Δ)

In this section, we are concerned with almost sure convergence of supermartingales (M_n) as $n \to \infty$. Naturally, martingales are a special case and submartingales can be handled through multiplying by -1. We'll need the following definition:

Definition 8.2.1 *Let $p \in [1, \infty)$. We say that a stochastic process (X_n) is bounded in L^p if*

$$\sup_{n \in \mathbb{N}} \mathbb{E}\left[|X_n|^p\right] < \infty.$$

Let (M_n) be a stochastic process and fix $a < b$. We define $U_N[a, b]$ to be the number of *upcrossings* made in the interval $[a, b]$ by $M_1, \ldots, M_N$. That is, $U_n[a, b]$ is the largest k such there exists

$$0 \leq s_1 < t_2 < \ldots < s_k < t_k \leq N \quad \text{ such that } \quad M_{s_i} \leq a, M_{t_i} > b \text{ for all } i = 1, \ldots, k.$$

Studying upcrossings is key to establishing almost sure convergence of supermartingales. To see why upcrossings are important, note that if $(c_n) \subseteq \mathbb{R}$ is a (deterministic) sequence and $c_n \to c$, for some $c \in \mathbb{R}$, then there is no interval $[a, b]$ $a < b$ such that $(c_n)_{n=1}^{\infty}$ makes infinitely many upcrossings of $[a, b]$; if there was then (c_n) would oscillate and couldn't converge.

Lemma 8.2.2 (Doob's Upcrossing Lemma) *Let M be a supermartingale. Then*

$$(b - a)\mathbb{E}[U_N[a, b]] \leq \mathbb{E}[(M_N - a)^-].$$

PROOF: Let $C_1 = \mathbb{1}\{M_0 < a\}$ and recursively define

$$C_n = \mathbb{1}\{C_{n-1} = 1, M_{n-1} \leq b\} + \mathbb{1}\{C_{n-1} = 0, M_{n-1} < a\}.$$

The behaviour of C_n is that, when X enters the region below a, C_n starts taking the value 1. It will continue to take the value 1 until M enters the region above b, at which point C_n will start taking the value 0. It will continue to take the value 0 until M enters the region below a, and so on. Hence,

$$(C \circ M)_N = \sum_{k=1}^{N} C_k(M_{k+1} - M_k) \geq (b - a)U_N[a, b] - (M_N - a)^-.$$

That is, each upcrossing of $[a, b]$ by M picks up at least $(b - a)$; the final term corresponds to an upcrossing that M might have started but not finished.

Note that C is previsible, bounded and non-negative. Hence, by Theorem 7.1.1 we have that $C \circ X$ is a supermartingale. Thus $\mathbb{E}[(C \circ M)_N] \leq 0$, which proves the given result. ∎

Note that $U_N[a, b]$ is an increasing function of N, and define $U_\infty[a, b]$ by

$$U_\infty[a, b](\omega) = \lim_{N \uparrow \infty} U_N[a, b](\omega).$$

With this definition, $U_\infty[a, b]$ could potentially be infinite, but we can prove that it is not.

Lemma 8.2.3 *Suppose M is a supermartingale and bounded in L^1. Then $P[U_\infty[a, b] = \infty] = 0$.*

Proof: From Lemma 8.2.2 we have

$$(b-a)\mathbb{E}[U_N[a,b]] \leq |a| + \sup_{n\in\mathbb{N}} \mathbb{E}|M_n| < \infty.$$

Hence, by the dominated convergence theorem we have

$$(b-a)\mathbb{E}[U_\infty[a,b]] \leq |a| + \sup_{n\in\mathbb{N}} \mathbb{E}|M_n| < \infty,$$

which implies that $\mathbb{P}[U_\infty[a,b] < \infty] = 1$. ■

Essentially, Lemma 8.2.3 says that the paths of M cannot oscillate indefinitely. This is the crucial ingredient of the martingale convergence theorem.

Theorem 8.2.4 (Martingale Convergence Theorem I) *Suppose M is a supermartingale bounded in L^1. Then the almost sure limit $M_\infty = \lim_{n\to\infty} M_n$ exists and $\mathbb{P}[|M_\infty| < \infty] = 1$.*

Proof: Define

$$\Lambda_{a,b} = \{\omega : \text{ for infinitely many } n, M_n(\omega) < a\} \cap \{\omega : \text{ for infinitely many } n, M_n(\omega) > b\}.$$

We observe that $\Lambda_{a,b} \subset \{U_\infty[a,b] = \infty\}$, which has probability 0 by Lemma 8.2.3. But since

$$\{\omega : M_n(\omega) \text{ does not converge to a limit in } [-\infty,\infty]\} = \bigcup_{a,b\in\mathbb{Q}} \Lambda_{a,b},$$

we have that

$$\mathbb{P}[M_n \text{ converges to some } M_\infty \in [-\infty,+\infty]] = 1$$

which proves the first part of the theorem.

($\star$) To prove the second part, we need to use a result from Lebesgue integration known as Fatou's Lemma. In our situation, Fatou's lemma implies that $\mathbb{E}[|M_\infty|] \leq \sup_n \mathbb{E}[|M_n|]$. Since we assume that (M_n) is bounded in L^1, we have $\mathbb{E}[|M_\infty|] \leq \sup_n \mathbb{E}[|M_n|] < \infty$ which implies that $\mathbb{P}[|M_\infty| < \infty] = 1$. ■

One useful note is that if M_n is a non-negative supermartingale then we have $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] \leq \mathbb{E}[M_0]$, so in this case M is automatically bounded in L^1.

Theorem 8.2.4 has one big disadvantage: it cannot tell us anything about the limit M_∞, except that it is finite. To gain more information about M_∞, we need an extra condition.

Corollary 8.2.5 (Martingale Convergence Theorem II) *In the setting of Theorem 8.2.4, suppose additionally that (M_n) is bounded in L^2. Then $M_n \to M_\infty$ in both L^1 and L^2, and*

$$\lim_{n\to\infty} \mathbb{E}[M_n] = \mathbb{E}[M_\infty], \qquad \lim_{n\to\infty} \mathrm{var}(M_n) \to \mathrm{var}(M_\infty).$$

The proof of Corollary 8.2.5 is outside of the scope of our course.

8.3 Long term behaviour of Galton-Watson processes (Δ)

The Borel-Cantelli lemmas and the martingale convergence theorem can tell us about the long term behaviour of stochastic processes. In this section we focus on the Galton-Watson process, which we introduced in Section 4.3.

Let us recall the notation from Section 4.3. Let X_i^n, where $n, i \geq 1$, be i.i.d. nonnegative integer-valued random variables with common distribution G. Define a sequence $(Z_n))$ by $Z_0 = 1$ and

$$Z_{n+1} = \begin{cases} X_1^{n+1} + \ldots + X_{Z_n}^{n+1}, & \text{if } Z_n > 0 \\ 0, & \text{if } Z_n = 0 \end{cases}$$

Then (Z_n) is the Galton-Watson process. We showed that, writing $\mu = \mathbb{E}[G]$,

$$M_n = \frac{Z_n}{\mu^n}$$

was a martingale.

We now look to describe the long term behaviour of the process Z_n, meaning that we want to know how Z_n behaves as $n \to \infty$. We'll consider three cases: $\mu < 1$, $\mu = 1$, and $\mu > 1$. Before we start, its important to note that, if $Z_N = 0$ for any $N \in \mathbb{N}$, then $Z_n = 0$ for all $n \geq N$. When this happens, it is said that the process *dies out*.

Lemma 8.3.1 *Suppose that $\mu < 1$. Then $\mathbb{P}[Z_n = 0 \text{ e.v.}] = 1$.*

PROOF: Since M_n is a martingale we have $\mathbb{E}[M_n] = \mathbb{E}[M_0] = 1$ for all n. Hence, $\mathbb{E}[Z_n] = \mu^n$. Note that

$$\mathbb{P}[Z_n \neq 0] = \sum_{k=1}^{\infty} \mathbb{P}[Z_n = k] \leq \sum_{k=0}^{\infty} k\mathbb{P}[Z_n = k] = \mathbb{E}[Z_n] = \mu^n.$$

Using fact, along with $\mu < 1$, we have $\sum_n \mathbb{P}[Z_n \neq 0] \leq \sum_n \mu^n < \infty$. Hence, the first Borel-Cantelli lemma gives us that $\mathbb{P}[Z_n \neq 0 \text{ i.o.}] = 0$, or in other words, $\mathbb{P}[Z_n = 0 \text{ e.v.}] = 1$. ■

The other two cases are more delicate and we'll need the martingale convergence theorem(s). We have that M_n is a martingale, and since $M_n \geq 0$ we have $\sup_n \mathbb{E}[|M_n|] = 1$, hence by the martingale convergence theorem (Theorem 8.2.4) we have that the almost sure limit

$$\lim_{n \to \infty} M_n = M_\infty$$

exists. Since $M_n \geq 0$, we have $M_\infty \in [0, \infty)$.

When $\mu = 1$, we get the same result as when $\mu < 1$, but we need a different method of proof.

Lemma 8.3.2 *Suppose that $\mu = 1$. Then $\mathbb{P}[Z_n = 0 \text{ e.v.}] = 1$.*

PROOF: In this case $M_n = Z_n$. So, we have $Z_n \to M_\infty$ almost surely. The offspring distribution is not deterministic, so for as long as $Z_n \neq 0$ the value of Z_n will continue to change over time. Each such change in value is of magnitude at least 1, hence the only way Z_n can converge is if Z_n is eventually zero. Therefore, since Z_n does converge, in this case we must have $\mathbb{P}[Z_n = 0 \text{ e.v.}] = 1$ and $M_\infty = 0$ ■

This leaves the case $\mu > 1$, which can be found as exercise 8.6. Let us write $\sigma^2 = \text{var}(G)$. The result is that:

Lemma 8.3.3 *Suppose that $\mu > 1$ and $\sigma^2 < \infty$. Then $\mathbb{P}[Z_n \to \infty] > 0$.*

The case of $\mu > 1$ and $\sigma^2 = \infty$ is more difficult and we don't study it in this course.

8.4 Exercises (Δ)

On Borel-Cantelli

8.1 Let $p > 0$. Let (U_n) be a sequence of i.i.d. uniform random variables on $(0, 1)$ and set $E_n = \{U < \frac{1}{n^p}\}$. For which values of p is it true that $\mathbb{P}[E_n \text{ i.o.}] = 1$, and for which values of p do we have $\mathbb{P}[E_n \text{ i.o.}] = 0$?

8.2 (a) Let (X_n) be independent random variables such that

$$X_n = \begin{cases} n^2 & \text{with probability } \frac{1}{n}, \\ 0 & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

(i) Show that $\mathbb{P}\left[X_n = n^2 \text{ infinitely often }\right] = 1$.

(ii) Show that $X_n \to 0$ in probability, but not almost surely or in L^1.

(b) Construct an example of a sequence of random variables that converges in L^1 but does not converge almost surely.

8.3 Let X_n be a sequence of independent random variables with $\mathbb{E}[X_n] = 0$. Suppose that, for some $M \in \mathbb{R}$, for all n we have $\mathrm{var}(X_n) \le M$. Set $S_n = \frac{1}{n}\sum_{i=1}^n X_i$. Show that, for any $\epsilon > 0$, $\mathbb{P}[|S_{n^2}| > \epsilon \text{ i.o.}] = 0$ and hence deduce that $S_{n^2} \to 0$ almost surely as $n \to \infty$.

8.4 Let (X_n) be an i.i.d. sequence of random variables. Let

$$A_m = \{X_m \ge X_n \text{ for all } n = 1, 2, \ldots, m\}.$$

When A_m occurs we say that the process (X_n) sets a record at time m. Show that $\mathbb{P}[A_m] = \frac{1}{m}$ and deduce that new records occur infinitely often i.e. $\mathbb{P}[A_m \text{ i.o.}] = 1$.

On martingale convergence

8.5 Recall the game of Roulette described in Section 7.2. Suppose that a gambler starts with £N (where $N > 0$) and bets £1 on each play until they run out of money. Show that, almost surely, the gambler eventually loses all of their money.

8.6 Let (Z_n) be the Galton-Watson process from Section 4.3, with the offspring distribution G. Suppose that $\mathbb{E}[G] = \mu$ and $\mathrm{var}(G) = \sigma^2 < \infty$. Set $M_n = \frac{Z_n}{\mu^n}$.

(a) Using (3.4), show that $\mathbb{E}[M_{n+1}^2] = \mathbb{E}[M_n^2] + \frac{\sigma^2}{\mu^{n+2}}$.

(b) Deduce that there exists a real valued random variable M_∞ such that $M_n \to M_\infty$ almost surely and in both L^1, L^2.

(c) Show that $\mathbb{P}[M_\infty > 0] > 0$, and hence that $\mathbb{P}[Z_n \to \infty] > 0$.

8.7 Let (B_n) and (M_n) be as in the urn process described in Section 4.2.

(a) Calculate the probability that the first k balls drawn are red and the next j balls drawn are black.

(b) Show that $\mathbb{P}\left[B_n = k\right] = \frac{1}{n+1}$ for all $0 \le k \le n$, and deduce that $\lim_{n\to\infty} \mathbb{P}\left[M_n \le p\right] = p$ for all $p \in [0, 1]$.

(c) Show that (B_n) converges almost surely as $n \to \infty$ and take a guess at the distribution of its limit.

Appendix A

Solutions to exercises

Chapter 1

1.1 At time 1 we would have $10(1+r)$ in cash and 5 units of stock. This gives the value of our assets at time 1 as $10(1+r)+5S_1$.

1.2 (a) Assuming we don't buy or sell anything at time 0, the value of our portfolio at time 1 will be $x(1+r)+yS_1$, where S_1 is as in the solution to **1.1**. Since we need to be certain of paying off our debt, we should assume a worst case scenario for S_1, that is $S_1 = d$. So, we are certain to pay off our debt if and only if

$$x(1+r)+ysd > K.$$

(b) Since certainty requires that we cover the worst case scenario of our stock performance, and $d < 1+r$, our best strategy is to exchange all our assets for cash at time $t=0$. This gives us

$$x+ys$$

in cash at time 0. At time 1 we would then have $(1+r)(x+ys)$ in cash at time 0 and could pay our debt providing that

$$(1+r)(x+ys) > K.$$

1.3 (a) We borrow s cash at time $t=0$ and use it to buy one stock. Then wait until time $t=1$, at which point we will a stock worth S_1, where S_1 is as in **1.1**. We sell our stock, which gives us S_1 in cash. Since $S_1 \geq sd > s(1+r)$, we repay or debt (plus interest) which costs us $s(1+r)$. We then have

$$S_1 - s(1+r) > 0$$

in cash. This is an arbitrage.

(b) We perform the same strategy as in (a), but with the roles of cash an stock swapped: We borrow 1 stock at time $t=0$ and then sell it for s in cash. Then wait until time $t=1$, at which point we will have sr in cash. Since the price of a stock is now S_1, where S_1 is as in **1.1**, and $S_1 \geq ds > (1+r)s$, we now buy one stock costing S_1 and repay the stockbroker, leaving us with

$$s(1+r) - S_1 > 0$$

in cash. This is an arbitrage.

1.4 We can calculate, using integration by parts, that

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^{\infty} x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}.$$

Similarly, we can calculate that

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_0^{\infty} x^2\lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2}$$

This gives that

$$\mathrm{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}.$$

1.5 We can calculate

$$\mathbb{E}[X_n] = 0 \cdot \mathbb{P}[X_n = 0] + n^2\mathbb{P}[X_n = n^2] = \frac{n^2}{n} = n \to \infty$$

as $n \to \infty$. Also,

$$\mathbb{P}[|X_n| > 0] = \mathbb{P}[X_n = n^2] = \frac{1}{n} \to 0$$

as $n \to \infty$.

1.6 We have

$$\begin{aligned}
\mathbb{P}[Y \le y] &= \mathbb{P}\left[\frac{X-\mu}{\sigma} \le y\right] \\
&= \mathbb{P}[X \le \mu + y\sigma] \\
&= \int_{-\infty}^{\mu+y\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.
\end{aligned}$$

We want to turn the above integral into the distribution function of the standard normal. To do so, substitute $z = \frac{x-\mu}{\sigma}$, and we obtain

$$\begin{aligned}
\mathbb{P}[Y \le y] &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2}} \sigma\, dz \\
&= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz.
\end{aligned}$$

Hence, Y has the distribution function of $N(0,1)$, which means $Y \sim N(0,1)$.

1.7 For $p \neq 1$,

$$\int_1^\infty x^{-p}\, dx = \left[\frac{x^{-p+1}}{-p+1}\right]_{x=1}^{\infty}$$

and we obtain a finite answer only if $-p+1 < 0$. When $p = -1$, we have

$$\int_1^\infty x^{-1}\, dx = [\log x]_{x=1}^{\infty} = \infty,$$

so we conclude that $\int_1^\infty x^{-p}\, dx$ is finite if and only if $p > 1$.

1.8 As $n \to \infty$,

- $e^{-n} \to 0$ because $e > 1$.
- $\sin\left(\frac{n\pi}{2}\right)$ oscillates through $0, 1, 0, -1$ and does not converge.
- $\frac{\cos(n\pi)}{n}$ converges to zero by the sandwich rule since $|\frac{\cos(n\pi)}{n}| \le \frac{1}{n} \to 0$.
- $\sum\limits_{i=1}^{n} 2^{-i} = 1 - 2^{-n}$ is a geometric series, and $1 - 2^{-n} \to 1 - 0 = 1$.
- $\sum\limits_{i=1}^{n} \frac{1}{i}$ tends to infinity, because

$$\overbrace{\frac{1}{2}} + \overbrace{\frac{1}{3} + \frac{1}{4}} + \overbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}} + \overbrace{\frac{1}{9} + \frac{1}{10} + + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16}} + \ldots.$$

each term contained in a $\overbrace{\cdot}$ is greater than or equal to $\frac{1}{2}$.

1.9 From the previous question we have $\sum_{n=1}^{\infty} n^{-2} < \infty$. Define $n_1 = 1$ and then

$$n_{r+1} = \inf\{k\,;\, n > n_r \text{ and } |x_k| <)(r+1)^{-2}\}.$$

Then for all r we have $|x_{n_r}| \le r^{-2}$. Hence,

$$\sum_{r=1}^{\infty} |x_{n_r}| \le \sum_{r=1}^{\infty} r^{-2} < \infty.$$

Chapter 2

2.1 The result of rolling a pair of dice can be written $\omega = (\omega_1, \omega_2)$ where $\omega_1, \omega_2 \in \{1,2,3,4,5,6\}$. So a suitable Ω would be

$$\Omega = \Big\{(\omega_1, \omega_2)\,;\, \omega_1, \omega_2 \in \{1,2,3,4,5,6\}\Big\}.$$

Of course other choices are possible too. Since our choice of Ω is finite, a suitable σ-field is $\mathcal{F} = \mathcal{P}(\Omega)$.

2.2 (a) Consider the case of $\mathcal{F}$. We need to check the three properties in the definition of a σ-field.

1. We have $\emptyset \in \mathcal{F}$ so the first property holds automatically.
2. For the second we check compliments: $\Omega \setminus \Omega = \emptyset$, $\Omega \setminus \{1\} = \{2,3\}$, $\Omega \setminus \{2,3\} = \{1\}$ and $\Omega \setminus \Omega = \emptyset$; in all cases we obtain an element of $\mathcal{F}$.
3. For the third, we check unions. We have $\{1\} \cup \{2,3\} = \Omega$. Including $\emptyset$ into a union doesn't change anything; including Ω into a union results in Ω. This covers all possible cases, in each case resulting in an element of $\mathcal{F}$.

So, $\mathcal{F}$ is a σ-field. Since $\mathcal{F}'$ is just $\mathcal{F}$ with the roles of 1 and 2 swapped, by symmetry $\mathcal{F}'$ is also a σ-field.

(b) We have $\{1\} \in \mathcal{F} \cup \mathcal{F}'$ and $\{2\} \in \mathcal{F} \cup \mathcal{F}'$, but $\{1\} \cup \{2\} = \{1,2\}$ and $\{1,2\} \notin \mathcal{F} \cup \mathcal{F}'$. Hence $\mathcal{F} \cup \mathcal{F}'$ is not closed under unions; it fails property 3 of the definition of a σ-field.

However, $\mathcal{F} \cap \mathcal{F}'$ is the intersection of two σ-fields, so is automatically itself a σ-field. (Alternatively, note that $\mathcal{F} \cap \mathcal{F}' = \{\emptyset, \Omega\}$, and check the definition.)

(c) X is not $\mathcal{F}$ measurable. To see this, note that $X^{-1}(1) = \{1,3\}$ which is not an element of $\mathcal{F}$. However, for any $I \subseteq \mathbb{R}$ we have

$$X^{-1}(I) = \begin{cases} \emptyset & \text{if } 1,2 \notin I \\ \{1,3\} & \text{if } 1 \in I \text{ but } 2 \notin I \\ \{2\} & \text{if } 1 \notin I \text{ but } 2 \in I \\ \Omega & \text{if } 1,2, \in I. \end{cases}$$

In all cases we have $X^{-1}(I) \in \mathcal{F}'$, so X is $\mathcal{F}'$ measurable.

2.3 (a) Let A_m be the event that the sequence (X_n) contains precisely m heads. Let $A_{m,k}$ be the event that we see precisely m heads during the first k tosses and, from then on, only tails. Then,

$$\begin{aligned}
\mathbb{P}[A_{m,k}] &= \mathbb{P}[m \text{ heads in } X_1, \ldots, X_k,\ \text{no heads in } X_{k+1}, X_{k+2}, \ldots] \\
&= \mathbb{P}[m \text{ heads in } X_1, \ldots, X_k,]\, \mathbb{P}[\text{ no heads in } X_{k+1}, X_{k+2}, \ldots] \\
&= \mathbb{P}[m \text{ heads in } X_1, \ldots, X_k,] \times 0 \\
&= 0.
\end{aligned}$$

If A_m occurs then precisely one of $A_{m,1}, A_{m,2}, \ldots$ occurs. Hence,

$$\mathbb{P}[A_m] = \sum_{k=0}^{\infty} \mathbb{P}[A_{m,k}] = 0$$

.

(b) Let A be the event that the sequence (X_n) contains finitely many heads. Then, if A occurs, precisely one of $A_1, A_2, \ldots$ occurs. Hence,

$$\mathbb{P}[A] = \sum_{m=0}^{\infty} \mathbb{P}[A_m] = 0.$$

That is, the probability of having only finite many heads is zero. Hence, almost surely, the sequence (X_n) contains infinitely many heads.

By symmetry (exchange the roles of heads and tails) almost surely the sequence (X_n) also contains infinitely many tails.

2.4 We have $X(TT) = 0$, $X(HT) = X(TH) = 1$ and $X(HH) = 2$. This gives

$$\sigma(X) = \Big\{\emptyset, \{TT\}, \{TH, HT\}, \{HH\}, \{TT, TH, HT\}, \{TT, HH\}, \{TH, HT, HH\}, \Omega\Big\}.$$

To work this out, you could note that $X^{-1}(i)$ must be in $\sigma(X)$ for $i = 0,1,2$, then also include $\emptyset$ and Ω, then keep adding unions and complements of the events until find you have added them all. (Of course, this only works because $\sigma(X)$ is finite!)

2.5 Since sums, divisions and products of random variables are random variables, X^2+1 is a random variable. Since X^2+1 is non-zero, we have that $\frac{1}{X^2+1}$ is a random variable, and using products again, $\frac{X}{X^2+1}$ is a random variable.

For $\sin(X)$, recall that for any $x\in\mathbb{R}$ we have

$$\sin(x)=\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}x^{2n+1}.$$

Since X is a random variable so is $\frac{(-1)^n}{(2n+1)!}X^{2n+1}$. Since limits of random variables are also random variables, so is

$$\sin X=\lim_{N\to\infty}\sum_{n=0}^{N}\frac{(-1)^n}{(2n+1)!}X^{2n+1}.$$

2.6 Since $X\geq 0$ we have $\mathbb{E}[X^p]=\mathbb{E}[|X|^p]$ for all p.

$$\mathbb{E}[X]=\int_1^\infty x^1 2x^{-3}\,dx=\frac{2}{3}\int_1^\infty x^{-2}\,dx=2\left[\frac{x^{-1}}{-1}\right]_1^\infty=2<\infty$$

so $X\in L^1$, but

$$\mathbb{E}[X^2]=\int_1^\infty x^2 2x^{-3}\,dx=2\int_1^\infty x^{-1}\,dx=2\left[\log x\right]_1^\infty=\infty$$

so $X\notin L^2$.

2.7 Let $1\leq p\leq q<\infty$ and $X\in L^q$. Then,

$$\begin{aligned}\mathbb{E}[|X|^p]&=\mathbb{E}\left[|X|^p\mathbb{1}\{|X|\leq 1\}+|X|^p\mathbb{1}\{|X|>1\}\right]\\&\leq\mathbb{E}\left[1+|X|^q\right]\\&=1+\mathbb{E}|X|^q]<\infty\end{aligned}$$

Here we use that, if $|X|\leq 1$ then $|X|^p\leq 1$, and if $|X|>1$ then since $p\leq q$ we have $|X|^p<|X|^q$. So $X\in L^p$.

2.8 Let us do the case $Y=cX$, where $c\neq 0$. Then

$$Y^{-1}(I)=\{\omega\in\Omega\,;\,cX(\omega)\in I\}=\{\omega\in\Omega\,;\,X(\omega)\in cI\}=X^{-1}(cI)$$

where $cI=\{cx\,;\,x\in I\}$. If I is an interval of $\mathbb{R}$ then cI is also an interval of $\mathbb{R}$. Hence, for any interval I we have $Y^{-1}(I)=X^{-1}(cI)\in\sigma(X)$. By definition of $\sigma(Y)$ this means that $\sigma(Y)\subseteq\sigma(X)$.

Similarly, for any interval I we have $X^{-1}(I)=Y^{-1}(\frac{1}{c}I)$, which implies that $\sigma(X)\subseteq\sigma(Y)$. Hence in fact $\sigma(Y)=\sigma(X)$.

The case of $c+X$ is similar; the same argument works with addition in place of multiplication i.e. use $c+I=\{c+x\,;\,c\in I\}$ and $-c+I$ defined similarly.

2.9 Since $X\geq 0$ we can apply Markov's inequality, from Lemma 2.4.4, which gives us that

$$\mathbb{P}[X\geq a]\leq\frac{1}{a}\mathbb{E}[X]=0$$

for any $a>0$. Hence,

$$\mathbb{P}[X>0]=\mathbb{P}\left[\bigcup_{n=1}^{\infty}\{X\geq\tfrac{1}{n}\}\right]\leq\sum_{n=1}^{\infty}\mathbb{P}[X\geq\tfrac{1}{n}]=0$$

and hence $\mathbb{P}[X>0]=0$.

Chapter 3

3.1 We have $S_2=X_1+X_2$ so

$$\begin{aligned}\mathbb{E}[S_2\,|\,\sigma(X_1)]&=\mathbb{E}[X_1\,|\,\sigma(X_1)]+\mathbb{E}[X_2\,|\,\sigma(X_1)]\\&=X_1+\mathbb{E}[X_2]\\&=X_1.\end{aligned}$$

Here, we use the linearity of conditional expectation, followed by the fact that X_1 is $\sigma(X_1)$ measurable and X_2 is independent of $\sigma(X_2)$ with $\mathbb{E}[X_2] = 0$.
Secondly, $S_2^2 = X_1^2 + 2X_1X_2 + X_2^2$ so

$$\begin{aligned}\mathbb{E}[S_2^2] &= \mathbb{E}[X_1^2 \,|\, \sigma(X_1)] + 2\mathbb{E}[X_1X_2 \,|\, \sigma(X_1)] + \mathbb{E}[X_2^2 \,|\, \sigma(X_1)] \\ &= X_1^2 + 2X_1\mathbb{E}[X_2 \,|\, \sigma(X_1)] + \mathbb{E}[X_2^2] \\ &= X_1^2 + 1\end{aligned}$$

Here, we again use linearity of conditional expectation to deduce the first line. To deduce the second line, we use that X_1^2 and X_1 are $\sigma(X_1)$ measurable (using the taking out what is known rule for the middle term), whereas X_2 is independent of $\sigma(X_1)$. The final line comes from $\mathbb{E}[X_2 \,|\, \sigma(X_1)] = 0$ (which we already knew from above) and that $\mathbb{E}[X_2^2] = 1$.

3.2 (a) If $X = Y$ almost surely then $(X - Y)^2 = 0$ almost surely, so $\mathbb{E}[(X - Y)^2] = 0$. Conversely, if we have that $\mathbb{E}[(X - Y)^2] = 0$, then since $(X - Y)^2 \geq 0$ we have that $(X - Y)^2 = 0$ almost surely, hence $X = Y$ almost surely.

(b) We can calculate

$$\begin{aligned}\mathbb{E}[(X - Y)^2] &= \mathbb{E}[X^2 - 2XY + Y^2] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[XY] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[\mathbb{E}[XY \,|\, \mathcal{G}]] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[Y\mathbb{E}[X \,|\, \mathcal{G}]] \\ &= 2\mathbb{E}[Y^2] - 2\mathbb{E}[Y^2] \\ &= 0.\end{aligned}$$

To deduce the second line we use linearity and that $\mathbb{E}[X^2] = \mathbb{E}[Y^2]$. The third line follows from the tower rule and the fourth then follows by taking out what is known (since Y is $\mathcal{G}$ measurable). Lastly, we use our hypothesis that $Y = \mathbb{E}[X \,|\, \mathcal{G}]$. We now have that $\mathbb{E}[(X - Y)^2] = 0$, so by part (a) we have $X = Y$ almost surely.

3.3 For $i \leq n$, we have $X_i \in m\mathcal{F}_n$, so $S_n = X_1 + X_2 + \ldots + X_n$ is also $\mathcal{F}_n$ measurable. For each i we have $|X_i| \leq 2$ so $|S_n| \leq 2n$, hence $S_n \in L^1$. Lastly,

$$\begin{aligned}\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] &= \mathbb{E}[X_1 + \ldots + X_n \,|\, \mathcal{F}_n] + \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] \\ &= (X_1 + \ldots + X_n) + \mathbb{E}[X_{n+1}] \\ &= S_n.\end{aligned}$$

Here, we use linearity of conditional expectation in the first line. To deduce the second, we use that $X_i \in m\mathcal{F}_n$ for $i \leq n$ and that X_{n+1} is independent of $\mathcal{F}_\backslash$ in the second. For the final line we note that $\mathbb{E}[X_{n+1}] = \frac{1}{3}2 + \frac{2}{3}(-1) = 0$.

3.4 (a) Since $X_i \in \mathcal{F}_n$ for all $i \leq n$, we have $M_n = X_1X_2\ldots X_n \in m\mathcal{F}_n$. Since $|X_i| \leq c$ for all i, we have $|M_n| \leq c^n < \infty$, so M_n is bounded and hence $M_n \in L^1$ for all n. Lastly,

$$\begin{aligned}\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] &= \mathbb{E}[X_1X_2\ldots X_nX_{n+1} \,|\, \mathcal{F}_n] \\ &= X_1\ldots X_n\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] \\ &= X_1\ldots X_n\mathbb{E}[X_{n+1}] \\ &= X_1\ldots X_n \\ &= M_n.\end{aligned}$$

Here, to deduce the second line we use that $X_i \in m\mathcal{F}_n$ for $i \leq n$. To deduce the third line we use that X_{n+1} is independent of $\mathcal{F}_n$ and then to deduce the forth line we use that $\mathbb{E}[X_{n+1}] = 1$.

(b) By definition of conditional expectation (Theorem 3.1.1), we have $M_n \in L^1$ and $M_n \in m\mathcal{G}_n$ for all n. It remains only to check that

$$\begin{aligned}\mathbb{E}[M_{n+1} \,|\, \mathcal{G}_n] &= \mathbb{E}[\mathbb{E}[Z \,|\, \mathcal{G}_{n+1}] \,|\, \mathcal{G}_n] \\ &= \mathbb{E}[Z \,|\, \mathcal{G}_n] \\ &= M_n.\end{aligned}$$

Here, to get from the second to third lines we use the tower property.

3.5 (a) If $m = n$ then $\mathbb{E}[M_n \,|\, \mathcal{F}_n] = M_n$ because M_n is $\mathcal{F}_n$ measurable. For $m > n$, we have

$$\mathbb{E}[M_m|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[M_m \,|\, \mathcal{F}_{m-1}] \,|\, \mathcal{F}_n] = \mathbb{E}[M_{m-1} \,|\, \mathcal{F}_n].$$

Here we use the tower property to deduce the first equality (note that $m-1 \leq n$ so $\mathcal{F}_{m-1} \supseteq \mathcal{F}_n$) and the fact that (M_n) is a martingale to deduce the second inequality (i.e. $\mathbb{E}[M_m \,|\, \mathcal{F}_{m-1}] = M_{m-1}$). Iterating, from $m, m-1 \ldots, n+1$ we obtain that

$$\mathbb{E}[M_m \,|\, \mathcal{F}_n] = \mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n]$$

and the martingale property then gives that this is equal to M_n, as required.

(b) If (M_n) is a submartingale then $\mathbb{E}[M_m \,|\, \mathcal{F}_n] \geq M_n$, whereas if (M_n) is a supermartingale then $\mathbb{E}[M_m \,|\, \mathcal{F}_n] \leq M_n$.

3.6 We have $(M_{n+1} - M_n)^2 = M_{n+1}^2 - 2M_{n+1}M_n + M_n^2$ so

$$\begin{aligned}
\mathbb{E}\left[(M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n\right] &= \mathbb{E}[M_{n+1}^2 \,|\, \mathcal{F}_n] - 2\mathbb{E}[M_{n+1}M_n \,|\, \mathcal{F}_n] + \mathbb{E}[M_n^2 \,|\, \mathcal{F}_n] \\
&= \mathbb{E}[M_{n+1}^2 \,|\, \mathcal{F}_n] - 2M_n\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] + M_n^2 \\
&= \mathbb{E}[M_{n+1}^2 \,|\, \mathcal{F}_n] - 2M_n^2 + M_n^2 \\
&= \mathbb{E}[M_{n+1}^2 \,|\, \mathcal{F}_n] - M_n^2
\end{aligned}$$

as required. Here, we use the taking out what is known run (since $M_n \in m\mathcal{F}_n$) and the martingale property of (M_n).

It follows that $\mathbb{E}[M_{n+1}^2 \,|\, \mathcal{F}_n] - M_n^2 = \mathbb{E}[(M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n] \geq 0$. We have $M_n^2 \in m\mathcal{F}_n$ and since $M_n \in L^2$ we have $M_n^2 \in L^1$. Hence (M_n^2) is a submartingale.

3.7 Using that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = aX_n + bX_{n-1}$ we can calculate

$$\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] = \mathbb{E}[\alpha X_{n+1} + X_n \,|\, \mathcal{F}_n] = \alpha(aX_n + bX_{n-1}) + X_n = (\alpha a + 1)X_n + \alpha b X_{n-1}.$$

We want this to be equal to S_n, and $S_n = \alpha X_n + X_{n-1}$. So we need

$$\alpha a + 1 = \alpha \qquad \text{and} \qquad \alpha b = 1.$$

Hence, $\alpha = \frac{1}{b}$ is our choice. It is then easy to check that

$$\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] = \tfrac{1}{b}(aX_n + bX_{n-1}) + X_n = (\tfrac{a}{b} + 1)X_n + X_{n-1} = \tfrac{1}{b}X_n + X_{n-1} = S_n$$

and thus S_n is a martingale, as required.

3.8 Note that

$$\mathbb{E}[X_1 \,|\, \sigma(S_n)] = \mathbb{E}[X_2 \,|\, \sigma(S_n)] = \ldots = \mathbb{E}[X_n \,|\, \sigma(S_n)]$$

is constant, by symmetry (i.e. permuting the $X_1, X_2, \ldots, X_n$ does not change the distribution of S_n). Hence, if we set $\mathbb{E}[X_i \,|\, \sigma(S_n)] = \alpha$ then

$$n\alpha = \sum_{i=1}^n \mathbb{E}[X_i \,|\, \sigma(S_n)] = \mathbb{E}\left[\sum_{i=1}^n X_i \,\middle|\, \sigma(S_n)\right] = \mathbb{E}[S_n \,|\, \sigma(S_n)] = S_n.$$

Here we use the linearity of conditional expectation and the fact that S_n is $\sigma(S_n)$ measurable. Hence,

$$\alpha = \mathbb{E}[X_1 \,|\, \sigma(S_n)] = \frac{S_n}{n}.$$

Chapter 4

4.1 Since $Z_n = e^{S_n}$ and $S_n \in m\mathcal{F}_n$, also $Z_n \in m\mathcal{F}_n$. Since $|S_n| \leq n$, we have $|e^{S_n}| \leq e^n < \infty$, so Z_n is bounded and hence $Z_n \in L^1$. Hence also $M_n \in m\mathcal{F}_n$ and $M_n \in L^1$.

Lastly,

$$\begin{aligned}
\mathbb{E}[e^{S_n}|\mathcal{F}_n] &= \mathbb{E}[e^{X_{n+1}}e^{S_n} \,|\, \mathcal{F}_n] \\
&= e^{S_n}\mathbb{E}[e^{X_{n+1}} \,|\, \mathcal{F}_n] \\
&= e^{S_n}\mathbb{E}[e^{X_{n+1}}]
\end{aligned}$$

$$= e^{S_n}\frac{e+\frac{1}{e}}{2}.$$

Here, we use the taking out what is known rule and the fact that X_{n+1} (and hence also $e^{X_{n+1}}$) is independent of $\mathcal{F}_n$. This gives us that

$$\mathbb{E}[M_{n+1}\,|\,\mathcal{F}_n] = M_n$$
$$\mathbb{E}[Z_{n+1}\,|\,\mathcal{F}_n] \geq Z_n$$

where to deduce the last line we use that $\frac{e+\frac{1}{e}}{2} > 1$.

4.2 Since $X_i \in \mathcal{F}_n$ for all $i \leq n$ we have $S_n \in m\mathcal{F}_n$ where $\mathcal{F}_n = \sigma(X_1,\ldots,X_n)$. Further, if we set $m = \max(|a|,|b|)$ then $|S_n| \leq m$ so $S_n \in L^1$.
We have

$$\begin{aligned}\mathbb{E}[S_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}[S_n + X_{n+1}\,|\,\mathcal{F}_n]\\ &= S_n + \mathbb{E}[X_{n+1}\,|\,\mathcal{F}_n]\\ &= S_n + \mathbb{E}[X_{n+1}]\\ &= S_n + ap_a - bp_b\end{aligned}$$

where $p_b = 1-p_a$. Therefore, S_n is a martingale if and only if $ap_a = bp_b$.

4.3 The natural filtration is $\mathcal{F}_n = \sigma(X_1,\ldots,X_n)$. We can write

$$S_{n+1} = \mathbb{1}\{S_n = 0\} + \mathbb{1}\{S_n \neq 0\}(S_n + X_{n+1}).$$

Hence, if $S_n \in m\mathcal{F}_n$ then $S_{n+1} \in \mathcal{F}_{n+1}$. Since $S_1 = X_1 \in \mathcal{F}_n$, a trivial induction shows that $S_n \in \mathcal{F}_n$ for all n. hence, $L_n \in m\mathcal{F}_n$ and also $S_n + L_n \in m\mathcal{F}_n$.
We have $-n \leq S_n - L_n \leq n+1$ (because the walk is reflected at zero and can increase by at most 1 in each time step) so $S_n - L_n$ is bounded and hence $S_n - L_n \in L^1$.
Lastly,

$$\begin{aligned}\mathbb{E}[S_{n+1}|\mathcal{F}_n] &= \mathbb{E}[\mathbb{1}\{S_n = 0\} + \mathbb{1}\{S_n \neq 0\}(S_n + X_{n+1})\,|\,\mathcal{F}_n]\\ &= \mathbb{1}\{S_n = 0\} + \mathbb{1}\{S_n \neq 0\}(S_n + \mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right])\\ &= \mathbb{1}\{S_n = 0\} + \mathbb{1}\{S_n \neq 0\}S_n\\ &= \mathbb{1}\{S_n = 0\} + S_n.\end{aligned}$$

Here we use linearity and the taking out what is known ($S_n \in m\mathcal{F}_n$) rule, as well as that X_{n+1} is independent of $\mathcal{F}_n$. Hence, using what we discovered above and taking out what is known, we have

$$\begin{aligned}\mathbb{E}[S_{n+1} - L_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}\left[S_n - \sum_{i=0}^{n}\mathbb{1}\{S_i = 0\}\,\middle|\,\mathcal{F}_n\right]\\ &= S_n + \mathbb{1}\{S_n = 0\} - \sum_{i=0}^{n}\mathbb{1}\{S_i = 0\}\\ &= S_n - \sum_{i=0}^{n-1}\mathbb{1}\{S_i = 0\}\\ &= S_n - L_n\end{aligned}$$

which means that $S_n - L_n$ is a martingale.

4.4 Let B_n be the total of balls added that were red, up to and including at time n. Then, the proportion of red balls in the urn just after the n^{th} step is

$$M_n = \frac{1+B_n}{n+3}.$$

Essentially the same argument as for the two colour urn of Section 4.2 shows that M_n is adapted to the natural filtration and also in L^1. We have

$$\begin{aligned}\mathbb{E}[M_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}\left[M_{n+1}\mathbb{1}\{n^{th}\text{ draw is red}\} + M_{n+1}\mathbb{1}\{n^{th}\text{ draw is not red}\}\,\middle|\,\mathcal{F}_n\right]\\ &= \mathbb{E}\left[\frac{2+B_n}{n+4}\mathbb{1}\{n^{th}\text{ draw is red}\} + \frac{1+B_n}{n+4}\mathbb{1}\{n^{th}\text{ draw is not red}\}\,\middle|\,\mathcal{F}_n\right]\end{aligned}$$

$$
\begin{aligned}
&= \frac{2+B_n}{n+4}\mathbb{E}\left[\mathbb{1}\{n^{th} \text{ draw is red}\}\,\Big|\,\mathcal{F}_n\right] + \frac{1+B_n}{n+4}\mathbb{E}\left[\mathbb{1}\{n^{th} \text{ draw is not red}\}\,\Big|\,\mathcal{F}_n\right] \\
&= \frac{2+B_n}{n+4}\frac{1+B_n}{n+3} + \frac{1+B_n}{n+4}\left(1-\frac{1+B_n}{n+3}\right) \\
&= \frac{(n+4)+B_n(n+4)}{(n+3)(n+4)} \\
&= M_n.
\end{aligned}
$$

Hence (M_n) is a martingale.

4.5 By the result of **3.6**, S_n^2 is a submartingale.

(i) $M_n = S_n^2 + n$ is $\mathcal{F}_n$ measurable and since $|S_n| \le n$ we have $|M_n| \le n^2 + n < \infty$, so $M_n \in L^1$. Further, using the submartingale property of S_n^2,

$$\mathbb{E}[S_{n+1}^2 + n + 1 \,|\, \mathcal{F}_n] = \mathbb{E}[S_{n+1}^2 \,|\, \mathcal{F}_n] + n + 1 \ge S_n^2 + n$$

so M_n is a submartingale. But $\mathbb{E}[M_1^2] = 2 \ne 0 = \mathbb{E}[M_0^2]$, so (by Lemma 3.3.6) we have that (M_n) is not a martingale.

(ii) $M_n = S_n^2 - n$ has $M_n \in L^1$ and $M_n \in m\mathcal{F}_n$ for essentially the same reasons as we used in case (i). However,

$$
\begin{aligned}
\mathbb{E}[S_{n+1}^2 \,|\, \mathcal{F}_n] &= \mathbb{E}[(X_{n+1} + S_n)^2 \,|\, \mathcal{F}_n] \\
&= \mathbb{E}[X_{n+1}^2 \,|\, \mathcal{F}_n] + 2\mathbb{E}[X_{n+1}S_n \,|\, \mathcal{F}_n] + \mathbb{E}[S_n^2 \,|\, \mathcal{F}_n] \\
&= 1 + 2S_n\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] + \mathbb{E}[S_n^2 \,|\, \mathcal{F}_n] \\
&= 1 + 2S_n\mathbb{E}[X_{n+1}] + \mathbb{E}[S_n^2 \,|\, \mathcal{F}_n] \\
&= 1 + \mathbb{E}[S_n^2 \,|\, \mathcal{F}_n] \\
&= 1 + S_n^2.
\end{aligned}
$$

Here, we use the taking out what is known rule (since $S_n \in m\mathcal{F}_n$) along with the fact that X_{n+1} is independent of $\mathcal{F}_n$ and $\mathbb{E}[X_{n+1}] = 0$. Therefore,

$$\mathbb{E}[S_{n+1}^2 - (n+1) \,|\, \mathcal{F}_n] = S_n^2 - n$$

so as (M_n) is a martingale, hence also a submartingale.

(iii) Since S_n is a martingale, we have

$$\mathbb{E}\left[\frac{S_{n+1}}{n+1}\,\Big|\,\mathcal{F}_n\right] = \frac{1}{n+1}\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \frac{S_n}{n+1} \ne \frac{S_n}{n}.$$

Hence $M_n = \frac{S_n}{n}$ is not a martingale.

4.6 We showed that $S_n^2 - n$ was a martingale in **4.5**. We try to mimic that calculation and find out what goes wrong in the cubic case. So, we look at

$$
\begin{aligned}
\mathbb{E}[(S_{n+1} - S_n)^3 \,|\, \mathcal{F}_n] &= \mathbb{E}[S_{n+1}^3 \,|\, \mathcal{F}_n] - 3S_n\mathbb{E}[S_{n+1}^2 \,|\, \mathcal{F}_n] + 3S_n^2\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] - \mathbb{E}[S_n^3 \,|\, \mathcal{F}_n] \\
&= \mathbb{E}[S_{n+1}^3 \,|\, \mathcal{F}_n] - 3S_n(S_n^2 + 1) + 3S_n^2 S_n - S_n^3 \\
&= \mathbb{E}[S_{n+1}^3 \,|\, \mathcal{F}_n] - S_n^3 - 3S_n.
\end{aligned}
$$

Here we use linearity, taking out what is known and that fact that $\mathbb{E}[S_{n+1}^2 \,|\, \mathcal{F}_n] = S_n^2 + 1$ (from **4.5**). However, also $S_{n+1} - S_n = X_{n+1}$, so

$$
\begin{aligned}
\mathbb{E}[(S_{n+1} - S_n)^3 \,|\, \mathcal{F}_n] &= \mathbb{E}[X_{n+1}^3 \,|\, \mathcal{F}_n] \\
&= \mathbb{E}[X_{n+1}^3] \\
&= 0.
\end{aligned}
$$

because X_{n+1} is independent of $\mathcal{F}_n$ and $X_{n+1}^3 = X_n$ (it takes values only 1 or -1) so $\mathbb{E}[X_{n+1}^3] = 0$. Putting these two calculations together, we have

$$\mathbb{E}[S_{n+1}^3 \,|\, \mathcal{F}_n] = S_n^3 + 3S_n. \tag{A.1}$$

Suppose (aiming for a contradiction) that there is a deterministic function $f:\mathbb{N}\to\mathbb{R}$ such that $S_n^3-f(n)$ is a martingale. Then

$$\mathbb{E}[S_{n+1}^3-f(n+1)\,|\,\mathcal{F}_n]=S_n^3-f(n).$$

Combining the above equation with (A.1) gives

$$f(n+1)-f(n)=3S_n$$

but this is impossible because S_n is not deterministic. Thus we reach a contradiction; no such function f exists.

Chapter 5

5.1 The value of the portfolio $(1,3)$ at time 1 will be

$$1(1+r)+3S_1=(1+r)+3S_1$$

where S_1 is a random variable with $\mathbb{P}[S_1=su]=p_u$ and $\mathbb{P}[S_1=sd]=p_d$.

5.2 (a) We need a portfolio $h=(x,y)$ such that $V_1^h=\Phi(S_1)=1$, so we need that

$$\begin{aligned}x(1+r)+ysu&=1\\x(1+r)+ysd&=1.\end{aligned}$$

Solving this pair of linear equations gives $x=\frac{1}{1+r}$ and $y=0$. So our replicating portfolio consists simply of $\frac{1}{1+r}$ cash and no stock.

Hence, the value of this contingent claim at time 0 is $x+sy=\frac{1}{1+r}$.

(b) Now we need that

$$\begin{aligned}x(1+r)+ysu&=3\\x(1+r)+ysd&=1.\end{aligned}$$

Which has solution $x=\frac{1}{1+r}\left(3-\frac{2su}{sy-sd}\right)$ and $y=\frac{2}{s(u-d)}$.

Hence, the value of this contingent claim at time 0 is $x+sy=\frac{1}{1+r}\left(3-\frac{su}{sy-sd}\right)+\frac{1}{u-d}$.

5.3 (a) If we buy two units of stock, at time 1, for a price K then our contingent claim is $\Phi(S_1)=2S_1-K$.

(b) In a European put option, with strike price $K\in(sd,su)$, we have the option to sell a single unit of stock for strike price K. It is advantageous to us to do so only if $K>S_1$, so

$$\Phi(S_1)=\begin{cases}0 & \text{if } S_1=su,\\ K-S_1 & \text{if } S_1=sd.\end{cases}=\max(K-S_1) \tag{A.2}$$

(c) We $S_1=su$ we sell a unit of stock for strike price K and otherwise do nothing. So our contingent claim is

$$\Phi(S_1)=\begin{cases}K-S_1 & \text{if } S_1=su,\\ 0 & \text{if } S_1=sd.\end{cases}$$

(d) Holding both the contracts in (b) and (c) at once, means that in both $S_1=su$ and $S-1=sd$, we end up selling a single unit of stock for a fixed price K. Our contingent claim for doing so is $\Phi(S_1)-K-S_1$.

5.4 (a) The contingent claims for the call and put options respectively are $\max(S_1-K,0)$ from (5.6) and $\max(K-S_1,0)$ from (A.2). Using the risk neutral valuation formula from Proposition 5.2.6 we have

$$\Pi_0^{call}=\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\max(S_1-K,0)\right],\quad \Pi_0^{put}=\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\max(K-S_1,0)\right].$$

(b) Hence,

$$\begin{aligned}\Pi_0^{call}-\Pi_0^{put}&=\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\max(S_1-K,0)-\max(K-S_1,0)\right]\\&=\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[\max(S_1-K,0)+\min(S_1-K,0)\right]\end{aligned}$$

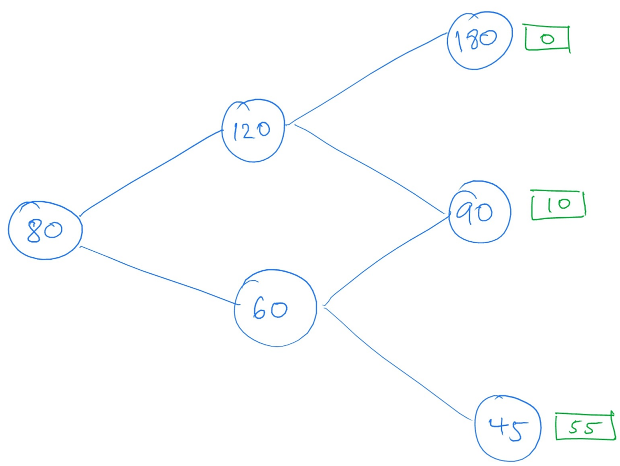
$$= \frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[S_1 - K\right]$$
$$= \frac{1}{1+r}\Big(\mathbb{E}^{\mathbb{Q}}\left[S_1\right] - K\Big)$$

By (5.4) (or you can use Proposition 5.2.6 again) we have $\frac{1}{1+r}\mathbb{E}^{\mathbb{Q}}\left[S_1\right] = S_0 = s$, so we obtain

$$\Pi_0^{call} - \Pi_0^{put} = s - \frac{K}{1+r}.$$

5.5 In the binomial model, the contingent claim of a European call option with strike price K is $\Phi(S_T) = \max(S_T - K)$.

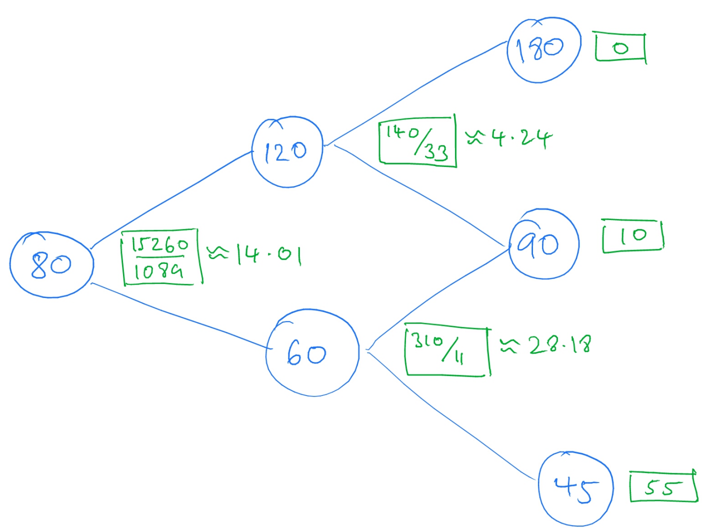
5.6 A tree-like diagram of the possible changes in stock price, with the value of the contingent claim $\Phi(S_T) = \max(K - S_1, 0)$ written on for the final time step looks like



By (5.3), the risk free probabilities in this case are

$$q_u = \frac{(1+0.1) - 0.75}{1.5 - 0.75} = \frac{8}{15}, \quad q_d = 1 - q_u = \frac{7}{15}.$$

Using Proposition 5.2.6 recursively, as in Section 5.6, we can put in the value of the European put option at all nodes back to time 0, giving



Hence, the value of our European put option at time 0 is 14.01 (to two decimal places).

5.7 We have that (S_t) is adapted (see Section 5.4) and since $S_t \in (d^t, u^t)$ we have also that $S_t \in L^1$. Hence, S_t is a martingale under $\mathbb{P}$ if and only if $\mathbb{E}^{\mathbb{P}}[S_{t+1} \,|\, \mathcal{F}_t] = S_t$. That is if

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[S_{t+1} \,|\, \mathcal{F}_t] &= \mathbb{E}[Z_{t+1} S_t \,|\, \mathcal{F}_t] \\ &= S_t \mathbb{E}^{\mathbb{P}}[Z \,|\, \mathcal{F}_t] \\ &= S_t \mathbb{E}^{\mathbb{P}}[Z]\end{aligned}$$

is equal to S_t. Hence, S_t is a martingale under $\mathbb{P}$ if and only if $\mathbb{E}^{\mathbb{P}}[Z_{t+1}] = up_u + dp_d = 1$.

Now consider $M_t = \log S_t$, and assume that $0 < d < u$ and $0 < s$, which implies $S_t > 0$. Since S_t is adapted, so is M_t. Since $S_t \in S_t \in (d^t, u^t)$ we have $M_t \in (t \log d, t \log u)$ so also $M_t \in L^1$. We have

$$M_t = \log\left(S_0 \prod_{i=1}^{t} Z_i\right) = \log(S_0) + \sum_{i=1}^{t} Z_i.$$

Hence,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}[M_{t+1} | \mathcal{F}_t] &= \log S_0 + \sum_{i=1}^{t} \log(Z_i) + \mathbb{E}[\log(Z_{t+1}) \,|\, \mathcal{F}_t] \\ &= M_t + \mathbb{E}^{\mathbb{P}}[\log Z_{t+1}] \\ &= M_t + p_u \log u + p_d \log d.\end{aligned}$$

Here we use taking out what is known, since S_0 and $Z_i \in m\mathcal{F}_t$ for $i \le t$, and also that Z_{t+1} is independent of $\mathcal{F}_t$. Hence, M_t is a martingale under $\mathbb{P}$ if and only if $p_u \log u + p_d \log d = 0$.

Chapter 6

6.1 Since $X_n \ge 0$ we have

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = \tfrac{1}{2} 2^{-n} = 2^{-(n+1)}$$

which tends to 0 as $n \to \infty$. Hence $X_n \to 0$ in L^1. Lemma 6.1.2 then implies that also $X_n \to 0$ in probability.

6.2 (a) We have $X_n \xrightarrow{L^1} X$ and hence

$$|\mathbb{E}[X_n] - \mathbb{E}[X]| = |\mathbb{E}[X_n - X]| \le \mathbb{E}[|X_n - X|] \to 0.$$

Here we use the linearity and absolute value properties of expectation. Hence, $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

(b) Let X_1 be a random variable such that $\mathbb{P}[X_1 = 1] = \mathbb{P}[X_1 = -1] = \frac{1}{2}$ and note that $\mathbb{E}[X_1] = 0$. Set $X_n = X_1$ for all $n \ge 2$, which implies that $\mathbb{E}[X_n] = 0 \to 0$ as $n \to \infty$. But $\mathbb{E}[|X_n - 0|] = \mathbb{E}[|X_n|] = 1$ so X_n does not converge to 0 in L^1.

6.3 We have $U^{-1} < 1$ so, for all $\omega \in \Omega$, $X_n(\omega) = U^{-n}(\omega) \to 0$ as $n \to \infty$. Hence $X_n(\omega) \to 0$ almost surely as $n \to \infty$. Further, $|X_n| \le 1$ for all $n \in \mathbb{N}$ and $\mathbb{E}[1] = 1 < \infty$, so the dominated convergence theorem applies and shows that $\mathbb{E}[X_n] \to \mathbb{E}[0] = 0$ as $n \to \infty$.

6.4 Let (X_n) be the sequence of random variables from **6.1**. Define $Y_n = X_1 + X_2 + \ldots + X_n$.

(a) For each $\omega \in \Omega$, we have $Y_{n+1}(\omega) = Y_n(\omega) + X_{n+1}(\omega)$. Hence, $(Y_n(\omega)$ is an increasing sequence. Since $X_n(\omega) \le 2^{-n}$ for all n we have

$$|Y_n(\omega)| \le 2^{-1} + 2^{-2} + \ldots + 2^{-n} \le \sum_{i=1}^{\infty} 2^{-1} = 1,$$

meaning that the sequence $Y_n(\omega)$ is bounded above by 1.

(b) Hence, since bounded increasing sequences converge, for any $\omega \in \Omega$ the sequence $Y_n(\omega)$ converges.

(c) The value of Y_1 is either 0 or $\frac{1}{2}$, both with probability $\frac{1}{2}$. The value of Y_2 is either $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, all with probability $\frac{1}{4}$. The value of Y_3 is either $0, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{7}{8}$ all with probability $\frac{1}{8}$.

(d) We can guess from (c) that the Y_n are becoming more and more evenly spread across $[0, 1]$, so we expect them to approach a uniform distribution as $n \to \infty$.

(e) We will use induction on n. Our inductive hypothesis is

(IH)$_n$: For all $k = 0, 1, \ldots, 2^n - 1$, it holds that $\mathbb{P}[Y_n = k2^{-n}] = 2^{-n}$.

In words, this says that Y_n is uniformly distributed on $\{k2^{-n}\,;\, k = 0, 1, \ldots, 2^n - 1\}$.

For the case $n = 1$, we have $Y_1 = X_1$ so $\mathbb{P}[Y_1 = 0] = \mathbb{P}[Y_1 = \frac{1}{2}] = \frac{1}{2}$, hence (IH)$_1$ holds.

Now, assume that (IH)$_n$ holds. We want to calculate $\mathbb{P}[Y_{n+1} = k2^{-(n+1)}$ for $k = 0, 1, \ldots, 2^{n+1} - 1$. We consider two cases, dependent on whether k is even or odd.

- If k is even the we can write $k = 2j$ for some $j = 0, \ldots, 2^n - 1$. Hence,

$$\begin{aligned}\mathbb{P}[Y_{n+1} = k2^{-(n+1)}] &= \mathbb{P}[Y_{n+1} = j2^{-n}]\\ &= \mathbb{P}[Y_n = j2^{-n} \text{ and } X_{n+1} = 0]\\ &= \mathbb{P}[Y_n = j2^{-n}]\mathbb{P}[X_{n+1} = 0]\\ &= 2^{-n}\tfrac{1}{2}\\ &= 2^{-(n+1)}.\end{aligned}$$

  Here we use that Y_n and X_{n+1} are independent, and use (IH)$_n$ to calculate $\mathbb{P}[Y_n = j2^{-n}]$.
- Alternatively, if k is odd then we can write $k = 2j + 1$ for some $j = 0, \ldots, 2^n - 1$. Hence,

$$\begin{aligned}\mathbb{P}[Y_{n+1} = k2^{-(n+1)}] &= \mathbb{P}[Y_{n+1} = j2^{-n} + 2^{-(n+1)}]\\ &= \mathbb{P}[Y_n = j2^{-n} \text{ and } X_{n+1} = 2^{-(n+1)}]\\ &= \mathbb{P}[Y_n = j2^{-n}]\mathbb{P}[X_{n+1} = 2^{-(n+1)}]\\ &= 2^{-n}\tfrac{1}{2}\\ &= 2^{-(n+1)}.\end{aligned}$$

  Here, again, we use that Y_n and X_{n+1} are independent, and use (IH)$_n$ to calculate $\mathbb{P}[Y_n = j2^{-n}]$.

In both cases we have shown that $\mathbb{P}[Y_{n+1} = k2^{-(n+1)}] = 2^{-(n+1)}$, hence (IH)$_{n+1}$ holds.

(f) From part (e), for any $a \in [0, 1]$ we have

$$(k-1)2^{-n} \leq a < k2^{-n} \quad \Rightarrow \quad \mathbb{P}[Y_n \leq a] = k2^{-n}.$$

Hence $a \leq \mathbb{P}[Y_n \leq a] \leq a + 2^{-n}$, which we write as

$$a \leq \mathbb{E}[\mathbb{1}\{Y_n \leq a\}] \leq a + 2^{-n}. \tag{A.3}$$

Since Y_n is increasing up to Y, we have that

- If $Y \leq a$ then $Y_n \leq a$ for all n.
- if $Y > a$ then for some $N \in \mathbb{N}$, for all $n \geq N$, $Y_n > a$.

Hence, $\mathbb{1}\{Y_n \leq a\} \to \mathbb{1}\{Y \leq a\}$ almost surely as $n \to \infty$. Since these random variables are bounded, we can apply the dominated converge theorem and deduce that

$$\mathbb{E}[\mathbb{1}\{Y_n \leq a\}] \to \mathbb{E}[\mathbb{1}\{Y \leq a\}$$

so taking limits in (A.3) gives us

$$a \leq \mathbb{E}[\mathbb{1}\{Y \leq a\}] \leq a$$

or, in other words, $a = \mathbb{P}[Y \leq a]$. Since we have shown this for arbitrary $a \in [0, 1]$, we can conclude that Y has distribution of a uniform random variable on $[0, 1]$.

6.5 By definition of convergence in probability, for any $a > 0$, for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that, for all $n \geq N$,

$$\mathbb{P}[|X_n - X| > a] < \epsilon \qquad \text{and} \qquad \mathbb{P}[|X_n - Y| > a] < \epsilon.$$

By the triangle inequality we have

$$\mathbb{P}[|X - Y| > 2a] = \mathbb{P}[|X - X_n + X_n - Y| > 2a] \leq \mathbb{P}[|X - X_n| + |X_n - Y| > 2a]. \tag{A.4}$$

If $|X - X_n| + |X_n - Y| > 2a$ then $|X - X_n| > a$ or $|X_n - Y| > a$ (or possibly both). Hence, continuing (A.4),

$$\mathbb{P}[|X - Y| > 2a] \leq \mathbb{P}[|X_n - X| > a] + \mathbb{P}[|X_n - Y| > a] \leq 2\epsilon.$$

Since this is true for any $\epsilon > 0$ and any $a > 0$, we have $\mathbb{P}[X = Y] = 1$.

6.6 Suppose (aiming for a contradiction) that there exists and random variable X such that $X_n \to X$ in probability. By the triangle inequality we have

$$|X_n - X_{n+1}| \leq |X_n - X| + |X - X_{n+1}|$$

Hence, if $|X_n - X_{n+1}| > 1$ then $|X_n - X| > \frac{1}{2}$ or $|X_{n+1} - X| > \frac{1}{2}$ (or both). Therefore,

$$\mathbb{P}\left[|X_n - X_{n+1}| > 1\right] \leq \mathbb{P}\left[|X_n - X| > \tfrac{1}{2}\right] + \mathbb{P}\left[|X_{n+1} - X| > \tfrac{1}{2}\right].$$

Since $X_n \to X$ in probability, the right hand side of the above tends to zero as $n \to \infty$. This implies that

$$\mathbb{P}\left[|X_n - X_{n+1}| > 1\right] \to 0$$

as $n \to \infty$. But, X_n and X_{n+1} are independent and $\mathbb{P}[X_n = 1, X_{n+1} = 0] = \frac{1}{4}$ so $\mathbb{P}[|X_n - X_{n+1}| > 1] \geq \frac{1}{4}$ for all n. Therefore we have a contradiction, so there is no X such that $X_n \to X$ in probability.

Hence, we can't have any X such that $X_n \to X$ almost surely or in L^1 (since it would imply convergence in probability).

Chapter 7

7.1 (a) Note that A_n is the indicator function of the event $\{S < n \leq T\} = \{S \leq n-1\} \setminus \{T \leq n-1\}$, which is an element of $\mathcal{F}_{n-1}$ since both S and T are stopping times. Hence, by Lemma 2.4.2 A_n is $\mathcal{F}_{n-1}$ measurable, so (A_n) is a previsible process.

(b) By definition of A_n, we have

$$(A \circ X)_n = \sum_{i=1}^{n} A_n(X_i - X_{i-1}) = \sum_{i=S\wedge n}^{T\wedge n} (X_i - X_{i-1}) = X_{T\wedge n} - X_{S\wedge n}.$$

(c) If (X_n) is a supermartingale then $(A \circ X)_n$ is also a supermartingale by Theorem 7.1.1. Hence $\mathbb{E}[(A \circ X)_n] \leq \mathbb{E}[(A \circ X)_0]$ for all n. In this case $(A \circ X)_0 = X_0 - X_0 = 0$ so

$$\mathbb{E}[X_{T\wedge n} - X_{S\wedge n}] \leq 0.$$

Hence $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_{S\wedge n}]$.

7.2 (a) Since T is a stopping time, and (from Section 4.1.2) $S_n - (p-q)n$ is a martingale, Theorem 7.3.4 tells us that $M_n = S_{T\wedge n} - (T \wedge n)(p-q)$ is a martingale. We have $\mathbb{E}[T] < \infty$ from Lemma 7.4.1 and

$$|M_{n+1} - M_n| \leq |S_{n+1} - S_n| + |p-q| \leq 1 + |p-q|$$

so condition (c) for the optional stopping theorem applies and $\mathbb{E}[M_T] = \mathbb{E}[M_0]$. That is,

$$\mathbb{E}[S_{T\wedge T} - (T \wedge T)(p-q)] = \mathbb{E}[S_{T\wedge 0} - (T\wedge 0)(p-q)] = \mathbb{E}[S_0 - 0(p-q)] = 0.$$

We thus have

$$\mathbb{E}[S_T] = (p-q)\mathbb{E}[T].$$

as required.

(b) We have

$$\begin{aligned}
\mathbb{E}[S_T] &= \mathbb{E}[S_T \mathbb{1}\{T = T_a\} + S_T \mathbb{1}\{T = T_b\}] \\
&= \mathbb{E}[a\mathbb{1}\{T = T_a\} + b\mathbb{1}\{T = T_b\}] \\
&= a\mathbb{P}[T = T_a] + b\mathbb{P}[T = T_b] \\
&= a\frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a} + b\frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}.
\end{aligned}$$

where in the last line we use equation (7.3). This gives us

$$\mathbb{E}[T] = \frac{1}{p-q}\left(a\frac{(q/p)^b - 1}{(q/p)^b - (q/p)^a} + b\frac{1 - (q/p)^a}{(q/p)^b - (q/p)^a}\right).$$

7.3 (a) We have $\{\max(S,T) \leq n\} = \{S \leq n\} \cup \{T \leq n\}$, which is an element of $\mathcal{F}_n$ since S and T are both stopping times. Hence $\max(S,T)$ is a stopping time.

(b) No. For example, for example if $S = 1$ (which is a stopping time) and T is any stopping time such that $T \geq 1$ then $\{T - S \leq n\} = \{T - 1 \leq n\} = \{T \leq n+1\}$, which is not, in general, $\mathcal{F}_n$ measurable.

7.4 (a) We have

$$T = \inf\{n > 0\,;\, B_n = 1\}.$$

Note that $T \geq n$ if and only if $B_i = 0$ for all $i = 1, 2, \ldots, n-1$. That is, if and only if we pick a red ball out of the urn at times $i - 1, 2, \ldots, n-1$. Hence,

$$\begin{aligned}\mathbb{P}[T \geq n] &= \mathbb{P}[B_1 = 0]\mathbb{P}[B_2 = 0]\ldots\mathbb{P}[B_{n-1} = 0]\\ &= \frac{1}{2}\frac{2}{3}\ldots\frac{n-2}{n-1}\\ &= \frac{1}{n-1}.\end{aligned}$$

Therefore, since $\mathbb{P}[T = \infty] \leq \mathbb{P}[T \geq n]$ for all n, we have $\mathbb{P}[T = \infty] = 0$ and $\mathbb{P}[T < \infty] = 1$.

(b) Since $M_n \in [0,1]$ the process (M_n) is bounded and we have shown that $\mathbb{P}[T < \infty] = 1$. Hence, we have condition (b) of the optional stopping theorem, so

$$\mathbb{E}[M_T] = \mathbb{E}[M_0] = \frac{1}{2}.$$

By definition of T we have $B_T = 1$. Hence $M_T = \frac{2}{T+2}$, so we obtain $\mathbb{E}[\frac{1}{T+2}] = \frac{1}{4}$.

7.5 (a) We use the filtration $\mathcal{F}_n = \sigma(N_i\,;\, i \leq n)$. We have $P_n \in m\mathcal{F}_n$ and since $0 \leq P_n \leq 1$ we have also that $P_n \in L^1$. Also,

$$\begin{aligned}\mathbb{E}[P_{n+1}\,|\,\mathcal{F}_n] &= \mathbb{E}[P_{n+1}\mathbb{1}\{\text{the } n^{th} \text{ ball was red}\}\,|\,\mathcal{F}_n] + \mathbb{E}[P_{n+1}\mathbb{1}\{\text{the } n^{th} \text{ ball was blue}\}\,|\,\mathcal{F}_n]\\ &= \mathbb{E}\left[\frac{N_n - 1}{2m-n-1}\mathbb{1}\{\text{the } n^{th} \text{ ball was red}\}\,\middle|\,\mathcal{F}_n\right] + \mathbb{E}\left[\frac{N_n}{2m-n-1}\mathbb{1}\{\text{the } n^{th} \text{ ball was blue}\}\,\middle|\,\mathcal{F}_n\right]\\ &= \frac{N_n - 1}{2m-n-1}\mathbb{E}[\mathbb{1}\{\text{the } n^{th} \text{ ball was red}\}\,|\,\mathcal{F}_n] + \frac{N_n}{2m-n-1}\mathbb{E}[\mathbb{1}\{\text{the } n^{th} \text{ ball was blue}\}\,|\,\mathcal{F}_n]\\ &= \frac{N_n - 1}{2m-n-1}\frac{N_n}{2m-n} + \frac{N_n}{2m-n-1}\frac{2m-n-N_n}{2m-n}\\ &= \frac{N_n}{2m-n}\\ &= P_n.\end{aligned}$$

Here we use taking out what is known (since $N_n \in m\mathcal{F}_n$), along with the definition of our urn process to calculate e.g. $\mathbb{E}[\mathbb{1}\{\text{the } n^{th} \text{ ball was red}\}\,|\,\mathcal{F}_n]$ as a function of N_n. Hence (P_n) is a martingale.

(b) Since $0 \leq P_n \leq 1$, the process (P_n) is a bounded martingale. The time T is bounded above by $2m$, hence condition (b) for the optional stopping theorem holds and $\mathbb{E}[P_T] = \mathbb{E}[P_0]$. Since $P_0 = \frac{1}{2}$ this gives us $\mathbb{E}[P_T] = \frac{1}{2}$. Hence,

$$\begin{aligned}\mathbb{P}[(T+1)^{st} \text{ ball is red}] &= \mathbb{P}[N_{T+1} = N_T + 1]\\ &= \sum_{i=1}^{2m-1}\mathbb{P}[N_{T+1} = N_T + 1\,|\,T = i]\mathbb{P}[T = i]\\ &= \sum_{i=1}^{2m-1}\frac{m-1}{2m-i}\mathbb{P}[T = i]\\ &= \mathbb{E}[P_T]\\ &= \frac{1}{2}.\end{aligned}$$

as required.

Chapter 8

8.1 We have $\mathbb{P}[E_n] = \frac{1}{n^p}$. So, $\sum_n \mathbb{P}[E_n] < \infty$ if $p > 1$ and $\sum_n \mathbb{P}[E_n] = \infty$ if $p \in (0,1]$. Hence, by the second Borel-Cantelli lemma, $\mathbb{P}[E_n \text{ i.o.}] = 0$ if $p > 1$ and $\mathbb{P}[E_n \text{ i.o.}] = 1$ if $p \in (0,1]$.

8.2 (a) Let (X_n) be independent random variables such that

$$X_n = \begin{cases} n^2 & \text{with probability } \frac{1}{n}, \\ 0 & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

(i) The X_n are independent and $\mathbb{P}[X_n = n^2] = \frac{1}{n}$ so as $\sum_n \mathbb{P}[X_n = n^2] = \infty$. Hence, by the second Borel-Cantelli lemma, $\mathbb{P}[X_n = n^2 \text{ i.o.}] = 1$. Similarly, $\sum_n \mathbb{P}[X_n = 0] = \sum_n (1 - \frac{1}{n}) \geq \sum_{n=2}^{\infty} \frac{1}{2} = \infty$, so $\mathbb{P}[X_n = 0 \text{ i.o.}] = 1$.

(ii) We have that, with probability 1, $X_n = 0$ for infinitely many n and $X_n = n^2$ for infinitely many n. Hence X_n oscillates and (with probability 1) does not converge. So X_n does not converge almost surely to 0.
For any $a > 0$ we have

$$\mathbb{P}[|X_n - 0| > a] = \mathbb{P}[X_n = n^2] = \frac{1}{n}$$

which tends to zero as $n \to \infty$. Hence $X_n \to 0$ in probability. Moreover,

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = n^2 \frac{1}{n} = n$$

which tends to ∞ as $n \to \infty$. Hence X_n does not converge to 0 in L^1.

(b) Define (X'_n) to be a sequence of independent random variables such that $\mathbb{P}[X'_n = 1] = \frac{1}{n}$ and $\mathbb{P}[X'_n = 0] = 1 - \frac{1}{n}$. Then, as in (a), we have $\mathbb{P}[X_n = 0 \text{ i.o.}] = 1$ and $\mathbb{P}[X_n = 1 \text{ i.o.}] = 1$. So, by the same argument as in (b), (X_n) does not converge almost surely. But

$$\mathbb{E}[|X_n - 0|] = \mathbb{E}[X_n] = \frac{1}{n}$$

which tends to zero as $n \to \infty$. Hence $X_n \to 0$ in L^1.

8.3 Let $\epsilon > 0$. Note that $\mathbb{E}[S_n] = 0$ by linearity and the fact that $\mathbb{E}[X_i] = 0$. Hence, from Markov's inequality (Lemma 2.4.4) we have

$$\mathbb{P}\left[|S_{n^2}| > \epsilon\right] = \mathbb{P}\left[\left|S_{n^2}^2\right| > \epsilon^2\right] \leq \frac{1}{\epsilon^2}\mathbb{E}\left[S_{n^2}^2\right] = \frac{1}{\epsilon^2}\operatorname{var}(S_n).$$

So, we need to calculate $\operatorname{var}(S_n)$.

$$\operatorname{var}(S_n) = \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(X_i) \leq \frac{1}{n^2}nM = \frac{M}{n}.$$

Note to swap $\sum$ and var we use the fact that the X_i are independent. Hence,

$$\operatorname{var}(S_{n^2}) \leq \frac{M}{n^2}$$

which means that

$$\mathbb{P}[|S_{n^2}| > \epsilon|] \leq \frac{M}{\epsilon^2 n^2}.$$

Hence, $\sum_n \mathbb{P}[|S_{n^2}| > \epsilon|] < \infty$ and the first Borel-Cantelli lemma implies that

$$\mathbb{P}\left[|S_{n^2}| > \epsilon \text{ i.o}\right] = 0.$$

which in turn means that $\mathbb{P}\left[|S_{n^2}| \leq \epsilon \text{ e.v}\right] = 1$.
That is, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that (with probability one) for all $n \geq N$, $|S_{n^2}| < \epsilon$. This means that $S_n \to 0$ almost surely.

8.4 For $n = 1, \ldots, m$ define

$$A_{m,n} = \{X_n \geq X_m \text{ for all } i = 1, \ldots, n-1, n+1, \ldots, m\}.$$

Then, by symmetry, $\mathbb{P}[A_{m,n}] = \mathbb{P}[A_{m,1}]$ for all $n = 1, \ldots, m$. Clearly we have $\cup_{n=1}^{m} A_{m,n} = \Omega$ and the events $A_{m,n}$ are disjoint for $n = 1, \ldots, m$, so we have

$$1 = \sum_{i=1}^{n} \mathbb{P}[A_{m,n}] = n\mathbb{P}[A_{m,1}].$$

Hence, $\mathbb{P}[A_m] = \frac{1}{m}$.

Clearly, we would now like to apply the second Borel-Cantelli lemma and deduce that $\mathbb{P}[A_m \text{ i.o.}] = 1$. But we cannot yet apply it, because we don't know if the events A_m are independent. So, we check

$$\begin{aligned}\mathbb{P}[A_m \cap A_{m+1} \cap \ldots \cap A_n] &= \mathbb{P}[A_m | A_{m+1} \cap A_{m-2} \cap \ldots \cap A_n]\mathbb{P}[A_{m+1} \cap A_{m+2} \cap \ldots \cap A_n] \\ &= \mathbb{P}[A_m]\mathbb{P}[A_{m+1} \cap A_{m+2} \cap \ldots \cap A_n]\end{aligned}$$

Here, the second line follows because A_m is independent of $A_{m+1} \cap A_{m+2} \cap \ldots \cap A_n$; this is because A_m depends only on the order of the first m variables, whereas A_{m+1}, A_{m+2} only depends on the first n variables through their maximum value (and the value of the max is independent of which position the max occurs at). So, by induction we deduce independence:

$$\mathbb{P}[A_m \cap A_{m+1} \cap \ldots \cap A_n] = P[A_m]\mathbb{P}[A_{m+1}] \ldots \mathbb{P}[A_n].$$

So, $\sum_m \mathbb{P}[A_m] = 1$ and by the second Borel-Cantelli lemma we obtain $\mathbb{P}[A_m \text{ i.o.}] = 1$.

8.5 On each play, our gambler increases his wealth, by 1, with probability $\frac{18}{37}$, or decreases his wealth, by 1, with probability $\frac{19}{37}$. We showed in Section 7.2 that their winnings M_n after n plays was a supermartingale. Since $M_n \geq 0$ we have $\mathbb{E}[|M_n|] = \mathbb{E}[M_n] \leq \mathbb{E}[M_0]$ so (M_n) is bounded in L^1. Hence, by the martingale convergence theorem, M_n is almost surely convergent.

Since M_n is integer valued, the only way that M_n can converge is if it becomes constant, eventually. That is, if there exists some (random) N such that M_n is constant for all $n \geq N$. But, since playing results in a win (an increase) or a loss (a decrease) the only way M_n can become eventually constant is if our gambler has lost all his money i.e. $M_n = 0$ eventually. So we conclude that $\mathbb{P}[M_n = 0 \text{ e.v.}] = 1$.

8.6 (a) We aim to use (3.4). With this in mind, since $M_n = \frac{Z_n}{\mu^n}$ we have

$$\frac{1}{\mu^{2(n+1)}}(Z_{n+1} - \mu Z_n)^2 = (M_{n+1} - M_n)^2. \tag{A.5}$$

Further,

$$Z_{n+1} - \mu Z_n = \sum_{i=1}^{Z_n} (X_i^{n+1} - \mu)$$

so it makes sense to define $Y_i = X_i^{n+1} - \mu$. Then $\mathbb{E}[Y_i] = 0$,

$$\mathbb{E}[Y_i^2] = \mathbb{E}[(X_i^{n+1} - \mu)^2] = \text{var}(X_i^{n+1}) = \sigma^2,$$

and $Y_1, Y_2, \ldots, Y_n$ are independent. Moreover, the Y_i are identically distributed and independent of $\mathcal{F}_n$. Hence, by taking out what is known ($Z_n \in m\mathcal{F}_n$) we have

$$\begin{aligned}\mathbb{E}\left[(Z_{n+1} - \mu Z_n)^2 \,|\, \mathcal{F}_n\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{Z_n} Y_i\right)^2 \,\middle|\, \mathcal{F}_n\right] \\ &= \sum_{i=1}^{Z_n} \mathbb{E}[Y_i^2 \,|\, \mathcal{F}_n] + \sum_{\substack{i,j=1 \\ i \neq j}}^{Z_n} \mathbb{E}[Y_i Y_j \,|\, \mathcal{F}_n] \\ &= \sum_{i=1}^{Z_n} \mathbb{E}[Y_i^2] + \sum_{\substack{i,j=1 \\ i \neq j}}^{Z_n} \mathbb{E}[Y_i]\mathbb{E}[Y_j] \\ &= Z_n \mathbb{E}[Y_1]^2 + 0 \\ &= Z_n \sigma^2.\end{aligned}$$

So, from (A.5) we obtain

$$\mathbb{E}[(M_{n+1}^2 - M_n)^2 \,|\, \mathcal{F}_n] = \frac{Z_n \sigma^2}{\mu^{2(n+1)}}.$$

which by (3.4) means that $\mathbb{E}[M_{n+1}^2 \,|\, \mathcal{F}_n] = M_n^2 + \frac{Z_n \sigma^2}{\mu^{2(n+1)}}$. Taking expectations, and using that $\mathbb{E}[Z_n] = \mu^n$, we obtain

$$\mathbb{E}[M_{n+1}^2] = \mathbb{E}[M_n^2] + \frac{\sigma^2}{\mu^{n+2}}.$$

(b) From (a), noting that $\mathbb{E}[M_0^2] = 1$, we have

$$\mathbb{E}[M_{n+1}^2] = 1 + \sum_{i=1}^{n} \frac{\sigma^2}{\mu^{i+2}} < 1 + \sum_{i=1}^{\infty} \frac{\sigma^2}{\mu^{i+2}} < \infty.$$

Here we use that $\mu > 1$ so a geometric series of $\frac{1}{\mu}$ converges. Therefore, (M_n) is bounded in L^2, and (recalling that M_n is a martingale!) the second martingale convergence theorem applies: there exists a real valued random variable M_∞ such that $M_n \to M_\infty$ almost surely and in both L^1, L^2.

(c) Furthermore, from the application of the second martingale convergence theorem in (b), we obtain that

$$\mathbb{E}[M_n] \to \mathbb{E}[M_\infty]$$

and since $\mathbb{E}[M_n] = 1$ this gives $\mathbb{E}[M_\infty] = 1$. Hence, $\mathbb{P}[M_\infty > 0] > 0$.

On the event that $M_\infty > 0$, we have $M_n = \frac{Z_n}{\mu^n} \to M_\infty > 0$. Since $\mu^n \to \infty$, this means that also $Z_n \to \infty$. Hence, we also have that $\mathbb{P}[Z_n \to \infty] > 0$.

8.7 (a) Since balls are drawn independently, the probability that the first k balls are black and the next j are red is

$$\frac{1}{2} \cdot \frac{2}{3} \cdot \ldots \cdot \frac{k-1}{k} \cdot \frac{k}{k+1} \cdot \frac{1}{k+2} \frac{2}{k+3} \cdot \ldots \cdot \frac{j}{k+j+1} = \frac{k!j!}{(k+j+2)!}$$

Drawing the same totals of red and black balls in a different order would have the same probability. The denominators in the above expression would come in a different order, but the overall value would stay the same.

(b) If $B_n = k$, then in the first n draws we have found k red balls. If we consider which k draws of the n draws gave red balls, we find $\binom{n}{k}$ possible combinations of them. By (a), each such combination has probability $\frac{k!(n-k)!}{(n+2)!}$ of occurring. Therefore,

$$\mathbb{P}[B_n = k] = \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}.$$

Since $M_n = \frac{B_n+1}{n+2}$, we have $M_n \le p$ if and only if $B_n \le p(n+2) - 1$. Hence,

$$\mathbb{P}[M_n \le p] = \sum_{k=1}^{\lfloor p(n+2)-1 \rfloor} \mathbb{P}[B_n = k] = \frac{\lfloor p(n+2) - 1 \rfloor}{n+1} \to p$$

as $n \to \infty$. (Recall that $\lfloor x \rfloor$ denotes x rounded down to the nearest integer.)

(c) Since (M_n) is a bounded martingale, the martingale convergence theorem tells us that it converges almost surely to a limit M_∞. In view of (b) we would expect (but we haven't proved!) that the limit would satisfy $\mathbb{P}[M_\infty \le p] = p$. So we expect M_∞ to have a uniform distribution on $(0, 1)$. (Try to prove it, if you want a challenge.)